European Commission

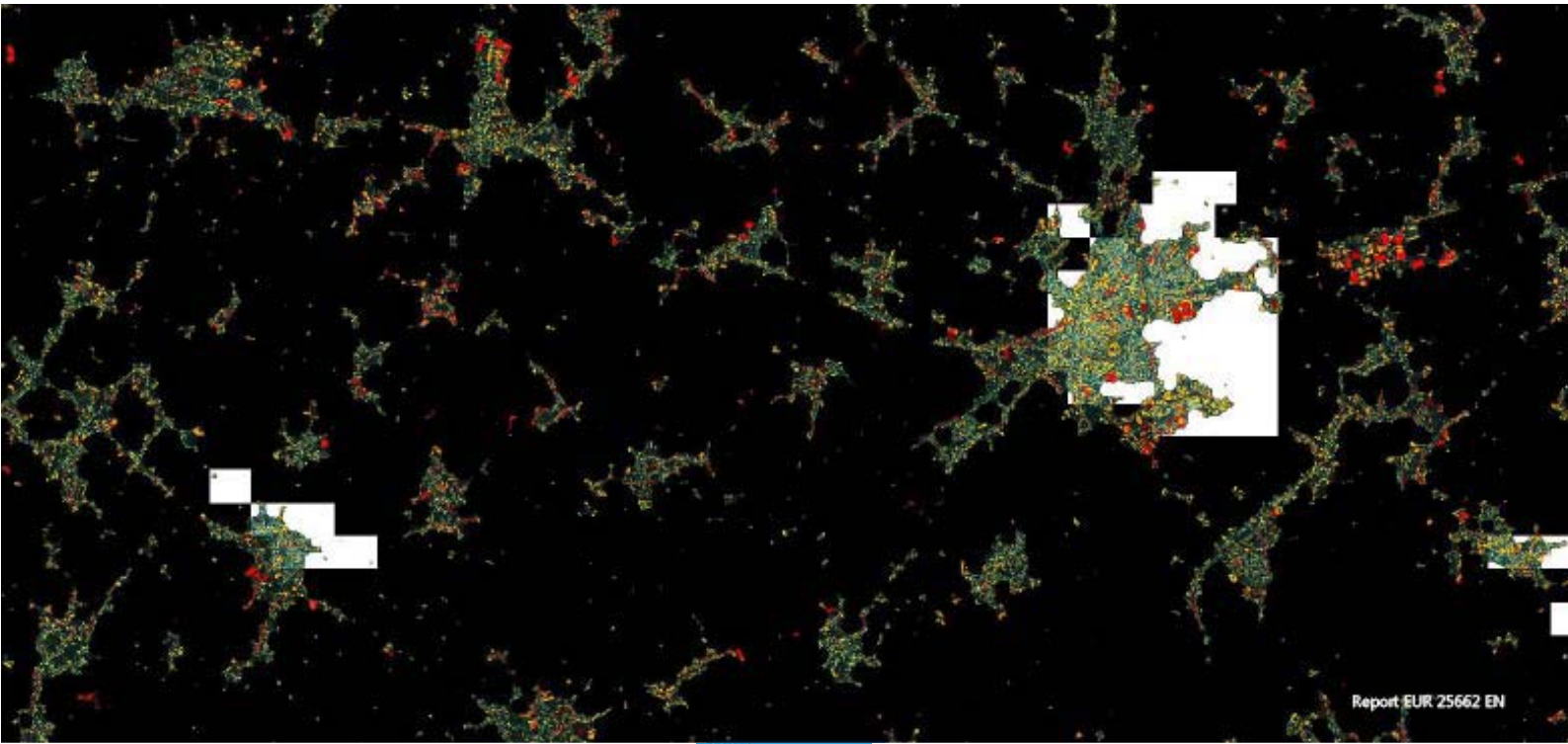
JRC SCIENTIFIC AND POLICY REPORTS

A Global Human Settlement Layer from Optical High Resolution Imagery

Concept and First Results

Martino Pesaresi, Xavier Blaes, Daniele Ehrlich, Stefano Ferri, Lionel Gueguen,, Fernand Haag, Matina Halkia, Johannes Heinzel, Mayeul Kauffmann, Thomas Kemper, Georgios K. Ouzounis, Marco Scavazzon, Pierre Soille, Vasileios Syrris, and Luigi Zanchetta

2012



Joint Research Centre

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Martino Pesaresi
Address: Joint Research Centre, Via Enrico Fermi 2749, TP267, 21027 Ispra (VA), Italy
E-mail: martino.pesaresi@jrc.ec.europa.eu
Tel.: +39 0332 78 9524
Fax: +39 0332 78 5154

http://ghslsys.jrc.ec.europa.eu/
http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC77925

EUR 25662 EN

ISBN 978-92-79-27988-1 (pdf)
ISBN 978-92-79-27989-8 (print)

ISSN 1831-9424 (online)
ISSN 1018-5593 (print)

doi:10.2788/73897

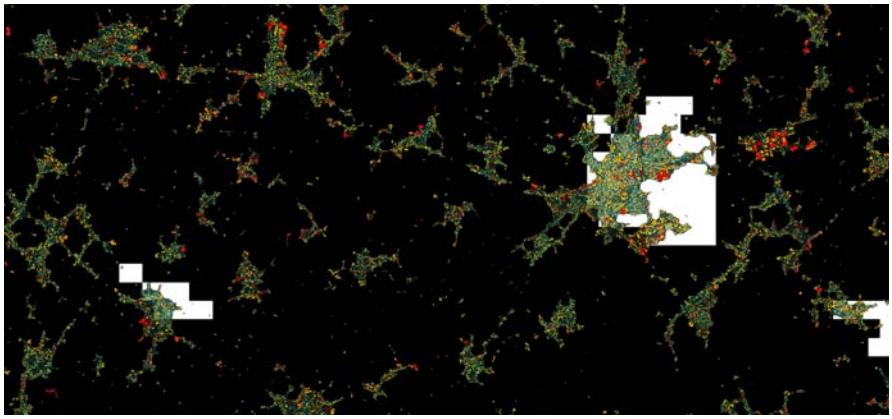
Luxembourg: Publications Office of the European Union, 2012

Printed in Italy

A Global Human Settlement Layer from optical high resolution imagery

Concept and first results

January 10, 2013



Martino Pesaresi, Xavier Blaes, Daniele Ehrlich, Stefano Ferri, Lionel Gueguen, Fernand Haag, Matina Halkia, Johannes Heinzel, Mayeul Kauffmann, Thomas Kemper, Georgios K. Ouzounis, Marco Scavazzon, Pierre Soille, Vasileios Syrris, and Luigi Zanchetta

European Commission, Joint Research Centre

Acknowledgements

This research was funded by the JRC Specific Programme of European Commission's Seventh Framework Programme for Research and Technological Development (FP7). It was undertaken under the work programme of the Geo-Spatial Information Analysis for Security and Stability action, Global Security and Crisis Management unit, Institute for the Protection and Security of the Citizen.

We are very grateful to Prof. Huadong Guo and Dr. Linlin Lu from the Chinese Academy of Science (CAS), Centre for Earth Observation and Digital Earth (CEODE), for their availability to process sample CBERS-2B imagery over China. Dr. Lionel Gueguen and Dr. Georgios K. Ouzounis contributed to this report whilst employed by the Joint Research Centre (JRC), European Commission, from 2009 to 2012.

Preface

This report proposes the Global Human Settlement Layer (GHSL) as a new way to map, analyse, and monitor human settlements. It reflects the work of the entire ISFEREA team over the last months. Although it has now reached some degree of maturity, it should not be regarded as a finished project. Instead it is work in progress that will be improved. This will include the technology needed to produce the layer as well as the of spatial coverage of the GHSL itself.

Contents

Acronyms

ACCA . . . . . Automatic Cloud Cover Assessment
AOI . . . . . Area Of Interest
AUC . . . . . Area Under the Curve
BIC . . . . . Bayesian Informatin Criterion
BU . . . . . Built-Up
BUREF . . . . . Built-Up REFerence
CBERS . . . . . China-Brazil Earth Resources Satellite
CID . . . . . Community Image Data portal
COPERNICUS European Programme for the establishment of a European capacity for Earth Observation (also called GMES)
CORINE. . . . . COoRdination of INformation on the Environment
CPU . . . . . Central Processing Unit
CSL . . . . . Characteristic Saliency Level
DAP . . . . . Differential Area/Attribute Profile
DB . . . . . DataBase
DEM . . . . . Digital Elevation Model
DMP . . . . . Derivative of the Morphological Profile
EC . . . . . European Commission
EER . . . . . Equal Error Rate
EPSG . . . . . European Petroleum Survey Group
EU . . . . . European Union
GB . . . . . GigaByte
GHSL . . . . . Global Human Settlement Layer
GIS . . . . . Geographic Information System
GLCM . . . . . Grey-Level Co-occurrence Matrix
GMES . . . . . Global Monitoring for Environment and Security (also called COPERNICUS)
GPU . . . . . Graphics Processing Unit
HIPARS . . . . High-performance Image Processing Algorithms for Remote Sensing
HPC . . . . . High Performance Computing
HR . . . . . High Resolution
HRC . . . . . High Resolution Camera
HSV . . . . . Hue-Saturation-Value
I2Q . . . . . Image Information Query
ID . . . . . IDentifyer

IDP . . . . . Internally Displaced Person
INPE . . . . . Instituto Nacional de Pesquisas Espaciais
INSPIRE . . . . . EU Directive to create a spatial data infrastructure
IQ . . . . . Image Query
JRC . . . . . Joint Research Centre
LSPOP . . . . . LandScan POPulation
LULC . . . . . Land Use Land Cover
MSE . . . . . Mean Squared Error
mtDAP . . . . . max-tree Differential Area/Attribute Profile
MUL . . . . . MULtispectral
NBU . . . . . Non-Built-Up
NDVI . . . . . Normalised Difference Vegetation Index
NIR . . . . . Near InfraRed
OGC . . . . . Open Geospatial Consortium
OSGeo . . . . . Open Source Geospatial Foundation
OSM . . . . . Open Street Map
PAN . . . . . PANchromatic
PSH . . . . . Pan-SHarpening
PSQL-PGIS . . . . . PostgreSQL-PostGIS
RAM . . . . . Random Access Memory
RGB . . . . . Red-Green-Blue
RMS . . . . . Root Mean Square
RMSE . . . . . Root Mean Squared Error
ROC . . . . . Receiver Operational Characteristics
RRM . . . . . Rapid Reaction Mechanism
RS . . . . . Remote Sensing
RSS . . . . . Rich Site Summary
SPOT . . . . . Satellite Probatoire d'Observation de la Terre
SRTM . . . . . Shuttle Radar Topography Mission
SSL . . . . . Soil Sealing Layer
TB . . . . . TeraByte
TMS . . . . . Tile Mapping Service
TOA . . . . . Top-Of-Atmosphere
UA . . . . . Urban Atlas
UN . . . . . United Nations
UN-HABITAT . . . . . United Nations Human Settlements Programme
UNHCR . . . . . Unite Nation's High Commission for Refugees
USGS . . . . . United States Geological Survey
UTM . . . . . Universal Transverse Mercator
VHR . . . . . Very High Resolution
WB . . . . . World Bank
WGS84 . . . . . World Geodetic System 1984
WMS . . . . . Web Mapping Service
WWF . . . . . World Wildflife Fund

List of Figures

List of Tables

Chapter 1
Introduction

This report describes the first operational test of the JRC image information query (I2Q) system to produce the first public release of the JRC global human settlement layer (GHSL).

The information query included both detection and an initial characterisation of built-up areas based on average size (scale) of built-up structures. These information where derived from a set of optical remotely-sensed imagery, which were stored in the JRC Community Image Data (CID) repository[1], or were processed in third-party servers.

In particular, this I2Q task involved the processing and evaluation of more than 15,000 satellite scenes having spatial resolutions in the range of 0.5–10 meters, and covering parts of Europe, South America, Asia and Africa, see Fig. 1.1. These image data were collected by a heterogeneous set of platforms including satellite SPOT (2 and 5), CBERS-2B, RapidEye (2 and 4), WorldView (1 and 2), GeoEye-1, QuickBird-2, IKONOS-2, and airborne sensors. The total surface mapped during the test is of more than 24 millions of square kilometres, involving the processing, indexing, and classification of around 4.07E+12 records (pixels) and their attributes (features) (Table 1.1).

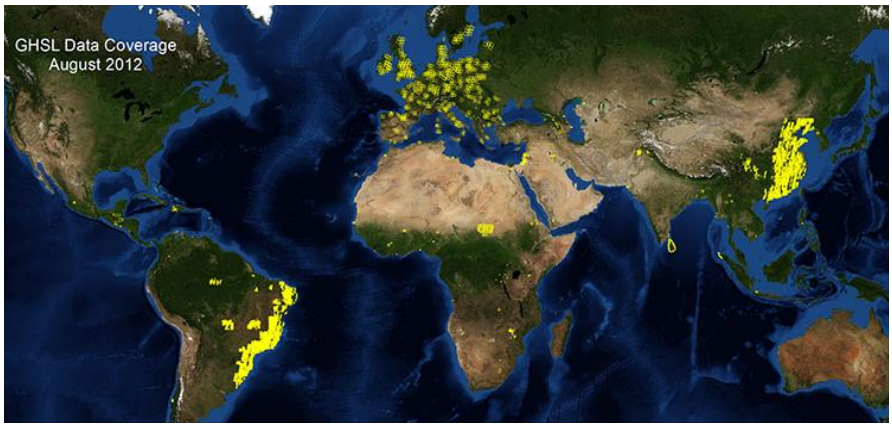


Fig. 1.1 Geographic distribution of the processed HR/VHR input images.

[1] URL: `http://cidportal.jrc.ec.europa.eu`

Table 1.1 Image datasets used in the experiment.

Platform name/version	Count of image ID	Sum of input surface (km^2)	Sum of input data volume (pix)
Aerial	6	9.21E+04	3.68E+11
CBERS-2B	9762	1.08E+07	1.72E+12
GeoEye-1	496	8.70E+04	9.42E+10
IKONOS-2	1421	3.47E+05	9.08E+10
QuickBird-2	1299	5.84E+05	3.26E+11
RapidEye-1	30	1.88E+04	4.44E+08
RapidEye-2	20	1.25E+04	2.96E+08
RapidEye-4	100	6.25E+04	1.48E+09
SPOT-2	1	5.01E+03	5.01E+07
SPOT-5	3168	1.22E+07	1.33E+12
WorldView-1	27	7.83E+03	2.40E+10
WorldView-2	402	1.41E+05	1.07E+11
Grand Total	16732	2.43E+07	4.07E+12

The purpose of the test reported here is two-fold. First, to study the feasibility and the system requirements necessary for the production of image-derived information layers supporting crisis management in realistic scenarios [Pes+10]. Second, to design and evaluate a specific image information retrieval task for the production of globally consistent outputs.

The keywords here are *realistic scenarios* and *globally consistent output*. In our definition, realistic scenarios include the presence of the following conditions:

1. *data volume:* large areas mapped (at least greater than half million of square kilometres);
2. *data complexity* including
   a. *detail:* high or very high spatial resolution image data input;
   b. *sensor heterogeneity:* availability of heterogeneous and arbitrary set of sensors varying both radiometric and spatial resolution of images;
   c. *quality heterogeneity:* different levels of pre-processing from geolocated raw data (level 2A) to pan sharpened ortho-rectified mosaics;
   d. *spatial inconsistency:* expected input RMS absolute positional accuracy ranging from 25 to 40 kilometres (CBERS-2B case);
   e. *seasonal arbitrariness:* arbitrary and not controllable set of the image data collection seasonality;
3. *suitable response time:* whole information retrieval efficiency of the order of 5,000 km^2/hour/CPU;
4. *low metadata quality:* uncontrollable level of documentation of the available data (including sensor characteristics, reflectance parameters, sun/satellite azimuth and elevation, and applied pre-processing steps) that can be also completely missing.

The term 'globally consistent output' refers to the capacity to produce standardised, comparable information across the globe. While doing this, we measure the consistency of the image information retrieval output against already existing global information layers, and we implemented the capacity to optimise it during the information production process.

In 2011 a first test of the JRC GHSL was performed, partially reported in [POG12], with the focus on the new multiscale morphological decomposition techniques applied during the workflow. The current experiment significantly improves the precedent one regarding i) the capacity to handle a much larger input data complexity and volume, ii) the capacity to measure the consistency with respect to available global information layers, and iii) the capacity to validate the GHSL output.

Specific segments of the feature extraction and image classification steps included in the current I2Q GHSL workflow were previously tested for automatic recognition and analysis of built-up areas. In particular, we used the `PanTex` image feature derived from anisotropic rotation-invariant grey level co-occurrence matrix (GLCM) textural contrast measurements and the characteristic-saliency-level (CSL) feature model [POG12]. The latter are based on Differential Morphological Profiles (DMP) [PB01] and derived morphological decomposition techniques [Dal+10; GSP; OPS12; OS12].

The capacity to discriminate built-up (BU) from non-built-up (NBU) areas of the PanTex features was previously assessed in a number of experiments including multi-temporal SPOT panchromatic data [PGK08; GSP11], and a set of 56 globally-representative VHR optical scenes covering large cities [Pes+11]. Strong correlation between PanTex image features and local density of building footprints was proven in [PHO11; OSP12] using cadastral data as reference. Similar techniques were also tested successfully using radar data input [Gam+08]. Moreover, the capacity to detect and describe the characteristics of the single built-up structures using DMP and DMP-derived image features were also previously tested in a number of experiments including optical data [Lu+12; BPK03; AA08], hyperspectral data [BPS05], and radar data [Mar+11]. These techniques demonstrated also effectiveness in characterisation of built-up areas as automatic detection of post-conflict damage assessment [PP08; Gue+12] post-earthquake damage assessment, [Li+10; Guo+09], destroyed buildings and rubble detection [OSP11] and more generally in image information mining tasks [Shy+06; GPS11]. The current experiment expands those findings to a much more general scenario.

Similar techniques integrating morphological and textural image features were also assessed in a first implementation of the JRC GHSL concept made in January 2011 that was using ENVISAT 75-m radar input imagery [TAP11b; TAP11a]. ENVISAT input datasets are not included in this test, because the realistic scenarios require HR input imagery with spatial resolution better than 10 meters.

While both the I2Q system design and the GHSL outputs where under assessment during the test reported here, the present paper will focus more on the description of the GHSL production workflow and the GHSL results. Only some general basic characteristics of the I2Q system and computational requirements will be shown.

This report is structured as follows. Chapter 2 presents the rationale of the GHSL production including application scenarios. The technical specifications are presented in Chapter 3. Chapter 4 describes the input data, the workflow as well as the system managing it. Pre-processing, feature extraction, and learning/classification are detailed in chapters 5, 6, and 7 respectively. Quality control is described in Chapter 8. Results are then discussed in Chapter 9. The report includes also two case studies on the development of the GHSL for Europe and Brazil, respectively. These are reported in the Chapters 10 and 11. The paper concludes in Chapter 12 with a summary, critical points, and the way forward.

Chapter 2
Rationale

The information on human settlements are crucial for a wide range of applications including emergency response, disaster risk reduction, population estimation/analysis, and urban/regional planning. Urbanisation pressure has an environmental impact, indicates population growth, and relates to risk and disaster vulnerability.

In 2011 the global population passed the mark of 7.0 billion and more than half of the population is living in urban areas. Between 2011 and 2050, the world population is expected to increase by 2.3 billion and the urban population to increase to 2.6 billion, passing from 3.6 billion in 2011 to 6.3 billion 2050 [UN 12]. The population growth in urban areas will be concentrated in the cities and towns of the less developed countries. Asia, in particular, is projected to see its urban population increase by 1.4 billion, Africa by 0.9 billion, and Latin America and the Caribbean by 0.2 billion. Population growth is therefore becoming largely an urban phenomenon concentrated in the developing world [Sat07]. The figures alone are alarming enough to understand that we are facing major challenges to manage the urban development in a sustainable way. A central issue in this respect is the availability of up-to-date information on the extent and quality of the urban settlement. In particular in less developed countries such information is largely unavailable. Cities are often growing at a pace that cannot be controlled by the local or regional mapping agencies.

Satellite imagery could provide information about the built environment worldwide, due to advances in computational and storage capacity, as well as data availability and cost. As demographic pressure increases exponentially at global level, our ability to monitor, quantify and characterise urbanisation processes around the world is becoming paramount. Despite this potential of remote sensing technologies, there are few global data sets that can be used to map the human settlements. Examples include the night-time lights of the world based on the DMSP-OLS sensor [Elv+01], MODIS based land use/land cover classifications [BB05; SFP10] and global population data sets like LandScan [Dob+00] or the gridded population of the world [CI12]. An overview, comparison and analysis of eight global data sets is provided in [Pot+09]. While these data sets are useful for global analysis, they have the tendency to under-represent small, scattered rural settlements due to their low spatial resolution between 500 and 2,000m. In addition, they represent a single snapshot in time that does not allow a regular monitoring. Or, if they are updated, like the LandScan data set, they are not directly comparable due to changing input sources. Although high resolution (HR, 1–10m spatial resolution) and even very high resolu-

tion (VHR, $\leq$ 1 m) data with an almost global coverage are available with different sensors (e.g., SPOT, CBERS, RapidEye, IKONOS, QuickBird, WorldView 1 and 2) no consistent global coverage of settlements derived from those datasets exists. Mapping and monitoring of urban areas at HR and VHR scales is mostly limited in terms of temporal and spatial coverage. They remain at the stage of case studies for individual or few cities and/or provide only a single time-step [Bau+10; NLM08; EB11]. The largest case study analyses 54 cities all around the world [Pes+11].

The lack of a consistent global layer with HR/VHR spatial resolution can be attributed to mainly two reasons. Firstly, the data availability of HR/VHR satellite data. Most HR/VHR satellite missions are operated on a commercial basis and consequently a global coverage is costly. The only relevant exception is the CBERS-2B platform releasing 2.5-m-resolution panchromatic data with a very open data sharing policy in Brazil. Secondly, to date no system has demonstrated the capacity to extract automatically global information layers about human settlement from HR/VHR satellite data with the necessary accuracy, and consequently so far only time-expensive manual or semi-automatic operational procedures were available.

The I2Q system design and GHSL production addresses the second point as the most important reason for the absence of a global human settlement layer made with HR/VHR spatial resolution imagery. Our vision is that an automatic image information retrieval system should contribute to reduce the amount of unexploited, but suitable image data stored in the archives of users and institutional stakeholders and covering the majority of populated places on the Earth surface. Such a system, if largely automatic, could be maintained by cost-effective budget not requiring large investments and operating costs for visual image interpreters. Process-on-demand services could be offered to users without image processing expertise to extract information related to human settlements from their own image data. The services could be linked to a collaborative environment, where the users agree on sharing of image-derived information. Inside this vision, we can imagine an incremental virtuous loop filling stepwise the gaps of a detailed digital representation of the global human settlement [Goo+12]. The above mentioned reasoning is explaining also some choices taken in the current experiment. For example, the emphasis on the ability to work with heterogeneous input image data including satellite and aerial sensors. This led to the decision to design an image information workflow robust enough to avoid crashing in presence of very low quality image data and metadata input. By tuning the system to this worst case input data/metadata scenario, we were able to automatically process on demand unknown quality input data, which is crucial in the light of the user scenarios of the system.

These scenarios are based to a large extend on the experience of the JRC in support to global crisis management. Since the aftermath of the Indian Ocean tsunami in 2004, the JRC offered operational support to European Commission services engaged in post-disaster damage and need assessment as well as disaster and crisis prevention through exploitation of remotely-sensed data interpretation. The JRC activities in the frame of the Instrument for Stability (IfS)[1] and its predecessor the Rapid Reaction Mechanism (RRM) included a number of disaster and crisis scenarios all around the globe, which contributed to the development of use scenarios for the GHSL. The GHSL concept and technical specifications were discussed also with JRC partners having global information needs on human settlements. In particular, intensive exchange of know-how regarding the possibility of automatic analysis of

[1] URL: `http://eeas.europa.eu/ifs/index_en.htm`

human settlement using remotely sensed data was done with International partners. There is collaboration with the World Bank, which is interested in globally consistent exposure mapping [Dei+11], with the Global Urban Observatory (GUO) of UN-Habitat[2], which monitoring of the Habitat Agenda and the Millennium Development Goals and activities pertaining to the production of reliable and up-to-date urban indicators at regional, country and city level, and with UNHCR's Field Information and Coordination Support Section (FICSS)[3] for the estimation of population in refugees and IDP camps [Cri99]. Use scenarios were also discussed with the JRC Global Disaster Alert and Coordination System (GDACS)[4], interested in a more accurate exposure information layer to use for their automatic impact and alert modelling. Finally, the broader societal impact of the GHSL concept both for technical-scientific and day-to-day users point of views, which were discussed in the frame of the new Digital Earth 2020 vision development [Goo+12; Cra+12].

Based on the above discussions, GHSL should be able to answer or contribute to answering the following questions:

- Is the often reported statistics that only 3% of the land masses are occupied by built up still valid?
- Shouldn't this statistics be updated given today's world population of 7 billion?
- What is the rate of sprawl/urbanisation in different regions of the world?
- How does the urbanisation process compare within a country and across countries?
- What are the new patterns of urban sprawl?
- How do the new urbanisation patterns change transport, energy consumption and sustainability?
- How does city extent and urban sprawl affect disaster risk and crisis management?
- Can we produce improved global fine scale physical exposure and population datasets to support global early warning systems and disaster loss models?

[2] URL: `http://www.unhabitat.org/categories.asp?catid=646`

[3] URL: `http://www.unhcr.org/pages/49c3646c4ca.html`

[4] URL: `http://gdacs.org/`

Chapter 3
GHSL definitions and specifications

3.1 GHSL definitions

The basic information contents of the current version of GHSL rely on the definition of built-up structures (buildings) and built-up areas: they are central for a quantitative description of human settlement based on HR and VHR remotely sensed imagery [PE09]. In the context of the GHSL, 'built-up areas' are the spatial generalisation of the notion of building and they are defined as follows:

Definition 1 A built-up area *is any given area or geographical space where buildings can be found.*

The size of the spatial domain used to generalise the buildings to built-up areas is classically a function of the *scale of representation* of a given geo-information product. The above definition can be translated directly in the visual reference data collection protocol used to verify the quality of the GHSL products at different scales.

In [ET12] the built-up area is defined as

> *'... a discrete area measurement that records the presence of buildings and the space in between buildings. The spaces in between buildings are defined by the spatial rule that defines the distance from the building. That distance is either a 1) buffer built around the building footprint or by 2) the grid cell size of the grid cell that intersect the buildings'.*

The approach of [ET12] can be made more consistent by linking the above notion of 'spatial rule' embedded in the built-up area to the concepts of *scale of representation* and *spatial generalisation* adopted in the cartographic representation of digital geo-information layers. Be $C_s(X)$ the cartographic representation of the information X at the scale S and be X the set of built-up structures represented in digital discrete entities at a given scale S, with $s \subset S$ expressing the size of the smallest spatial detail in $C_s(X)$, and assuming that if $s' > s$, then $C_{s'}(X)$ is the generalised (broader scale) version of $C_s(X)$. The $C_s(X)$ can be generalised to $C_{s'}(X)$ with $s' > s$ adopting three different approaches namely: 'inclusive', 'exclusive', and 'by sum'.

Be D_s any arbitrary spatial domain defined at the scale s of $C_s(X)$. The 'inclusive' notion of built up area is based on the following question: *'does the domain D_s hit the set X'?*. This can be formalised as $C_s(X) = \{x \mid D_s \cap X \neq \oslash\}$ that is the formula of set dilation transform known in mathematical morphology

[Ser82; Soi03]. The 'exclusive' notion of built up area is based on the following question: *'does the domain D_s fit the set X'?*. This can be formalised as $C_s(X) = \{x \mid D_s \subseteq X\}$ that is the formula of the morphological erosion transform. The 'by sum' notion of built-up area is based on the following question: *'how much the D_s overlaps the set X?'* that can be described as the convolution of X by the D_s kernel such that $C_s(X) = \{x \mid D_s * X\}$. It is worth noting that only aggregation or generalisation operations are allowed in this paradigm, starting from the finest scale available in the input source. Specifically, in this study this parameter is dictated by the spatial resolution of the input imagery used to collect the information about the presence of built-up areas, and the image features used for image information extraction and classification.

Independently from the strategy adopted for generalisation of built-up structures to built-up areas, a definition of built-up structure should be provided: the working definition of built-up structure (building) used in the frame of the GHSL production is as follows:

Definition 2 GHSL Built-up structures *are enclosed constructions above ground which are intended or used for the shelter of humans, animals, things or for the production of economic goods and that refer to any structure constructed or erected on its site.*

This working definition is adapted from the data specification on buildings delivered by the Infrastructure for Spatial Information in Europe (INSPIRE)[1] that is defined as follows:

> 'Buildings are enclosed constructions above **and/or underground**, which are intended or used for the shelter of humans, animals, things or for the production of economic goods and that refer to any structure **permanently** constructed or erected on its site.'

In the above definition, the text in bold differentiate the GHSL and the INSPIRE specifications of the notion of building. The INSPIRE concept of 'building' was adapted taking into account the specific GHSL constraints and user requirements. In particular, by contrast to the INSPIRE definition, the GHSL definition does not include underground building notion for obvious limitations of the considered input data.

Moreover, GHSL notion does not impose the permanency of the built-up structure on the site as instead INSPIRE does, following the classical topographic mapping tradition. The GHSL notion of built-up structure is more inclusive, accepting to describe also structures belonging to temporary human settlements as, for example, refugee or internal displaced people (IDP) camps. In 2011, there were an estimated 26.4 million people displaced internally by conflict[2] and 15.2 million refugees around the world, including 4.8 million Palestinian refugees[3]. These numbers don't include IDPs related to natural disasters and other forced migration

[1] INSPIRE Infrastructure for Spatial Information in Europe, 'D2.8.III.2 Data Specification on Building - Draft Guidelines', INSPIRE Thematic Working Group Building 2012 URL: `http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_BU_v2.0.pdf`

[2] URL: `http://www.internal-displacement.org/publications/global-overview-2011`

[3] UNHCR Global Trends 2011, URL: `http://www.unhcr.org/4fd6f87f9.html`

issues. The total number of refugees and IDPs was estimated in 2009 as 67 million of people[4].

Finally, in a different way than INSPIRE, the GHSL repository concept includes also 'slum' or informal settlement concepts: the area of a city characterised by substandard housing and poverty and lacking access to water and sanitation as well as security of tenure. These areas are often referred to as slums, shanty town, squatter settlement, favelas, barrios, or similar. According to the United Nations, due to rising population and the rise especially in urban populations, the number of slum dwellers is rising. One billion people worldwide live in slums and the figure is projected to grow to 2 billion by 2030[5].

The study of temporary and informal human settlements are important for addressing risk and vulnerability issues related to crisis management and disaster mitigation operations. For these reasons they are included in the GHSL specifications.

It is worth noting that the GHSL definition is only partially fitting with other similar available definitions already popular in the remote sensing community as the USGS *'urban or built-up areas'*[6], *'Impervious Surfaces'* [LW06], *'Urban Soil Sealing'*[7], CORINE *'Artificial Surfaces'*[8] and similar ones. Compared to these land-use/land-cover (LULC) definitions, the GHSL classification schema is more general not assuming any embedded urban/rural dichotomy (built-up structures are mapped independently if they are falling in any 'rural' or 'urban' area definitions) and more focused on quantitative support to crisis management, risk and disaster mitigation activities requiring detailed mapping of buildings, population and their vulnerabilities with a multi-scale approach. Furthermore, the GHSL classification schema including the simplification and lowering down of the embedded abstraction was designed in order to facilitate the semantic interoperability and then multi-disciplinary across-application sharing of data and results. This includes the sharing of data between different agencies (UN, WB, EC) working in similar areas but not necessarily sharing exactly the same abstract definitions[PE09].

3.2 GHSL technical specifications

This section includes a brief description of the GHSL technical specifications regarding format, scale, information production, and quality control. These specifications summarise the GHSL production guidelines.

---

[4] World Savvy Monitor, 2009, URL: `http://worldsavvy.org/monitor/index.php?option=com_content&view=article&id=441&Itemid=847`

[5] 'Slum Dwellers to double by 2030', UN-HABITAT report, April 2007. URL: `http://www.unhabitat.org/downloads/docs/4631_46759_GC%2021%20Slum%20dwellers%20to%20double.pdf`

[6] URL: `http://landcover.usgs.gov/urban/umap/htmls/defs.php`

[7] URL: `http://www.eea.europa.eu/articles/urban-soil-sealing-in-europe`

[8] URL: `http://www.eea.europa.eu/publications/COR0-landcover`

3.2.1 Format

The native storage format and structure of the GHSL information is tile-based: the basic spatial unit is a surface tile that is the representation of a given portion of the earth surface with a given size and projection. The tile entity is organised in a hierarchical multi-scale structure. The GHSL information is released trough standard protocols defined by the Open Geospatial Consortium (OGC). In particular, Web Map Service (WMS) and Tile Map Service (TMS) platforms release the GHSL product as output of a specific query to a spatial database (DB), including time and dynamic information queries. While satellite-derived image features are stored and managed in the local (UTM) metric projection, the global mosaic of the GHSL adopts a global metric projection, the Spherical Mercator (EPSG:900913) with WGS84 Datum.

3.2.2 Scale

GHSL information is built and provided in three nominal scales of reference, namely `local`, `regional` and `global` scales. They correspond to specific parameters regarding the TMS zoom level, the spatial unit of reference and the tolerance admitted in the geo-coding of the information. Table 3.1 describes the relation between these parameters in the design of the GHSL product. The output and the evaluation protocols are at this stage developed only at the regional and global scales, because only a minority of datasets available is fitting the local-scale 1:10K quality specifications.

Table 3.1 GHSL scales, spatial units, and tolerances.

Nominal Scale	Ratio	TMS zoom	Spatial Unit	Spatial RMS
local	1:10K	14	10m	5m
regional	1:50K	11	50m	25m
global	1:500K	8	500m	250m

3.2.3 Information contents

The I2Q system extracts two basic information layers from the imagery: namely the built up area presence BU_{area} and the built-up scale BU_{scale} for a given GHSL scale. The BU_{area} is a continuous value that is thresholded to provide binary information on the presence of buildings. The BU_{scale} provides a continuous value for the average scale (size) of buildings inside the given pixel. From this information the following GHSL variables are derived:

- $BuiltUpSurface$ expressed in m^2. It is the total surface of built-up area in the specific spatial unit calculated as $\sum BU_{\text{area}}$;
- $BuiltUpPercent$ percent of built-up surface in the specific spatial unit. It is calculated as $\sum BU_{\text{area}}/TileSurface$;

- AverageSurfaceOfBuildings expressed in m^2 —average size of buildings expressed as average surface of building footprints candidates in the specific spatial unit— $\sum(BuiltUpPercent * BU_{\text{scale}})$;
- NumberOfBuildings n —number of built-up structures estimated in the specific spatial unit— $\sum(BuiltUpSurface/AverageSurfaceOfBuildings)$.

It is worth noting that because of the hierarchical characteristics of the TMS structure used to store the final geo-information layer, the information located in the broader scales can be calculated from aggregation of the same information available in more detailed scales. For this information the general hierarchical relation $\sum \texttt{local} \Rightarrow \sum \texttt{regional} \Rightarrow \texttt{global}$ by aggregation (sum, average) is valid. Moreover, the spatial hierarchy is not only reflected in the aggregation rule, but also in the number of available descriptors. The number of abstract semantic layers available is increasing by increasing level of spatial generalisation by passing from local to global scales. Table 3.2 summarises the number of descriptors available in this version of the GHSL at the different scales.

3.2.4 Quality

For a system that is designed to work with a heterogeneous set of input data, quality control is essential. GHSL information layers are produced with various satellite image data, having different sensor and platform characteristics, and different pre-processing standards including different geo-coding and ortho-rectification protocols. Consequently, geo-coding quality of the input image data has to be checked and the derived image information will feed the GHSL scale corresponding to the assessed spatial tolerance. The input image spatial tolerance admitted is always less than half of the GHSL spatial unit output.

The quality of the product is tested by applying a public and reproducible protocol including i) visual image reference data collection and ii) systematic statistical distance measurements with respect to known reference layers available in any scale (local, regional, continental). The quality measures are embedded in the output metadata.

Table 3.2 GHSL descriptors and scales.

Descriptor	Local 1:10K	Regional 1:50K	Global 1:500K
TileSurface	X	X	X
BuiltUpSurface	X	X	X
BuiltUpPercent	X	X	X
AverageSurfaceOfBuildings	-	X	X
NumberOfBuildings	-	-	X

3.3 Computational requirements

The main workload in the production of the GHSL information layers is encountered at the level of feature extraction and in particular in the computation of building footprint candidates using the max-tree Differential Attribute Profile (mtDAP) protocol and the Characteristic Saliency Level (CSL) model (see Sec. 6). This section presents some statistics on their computation with reference to an experiment on sample image tiles.

The experiment was implemented on two modules of a High Performance Computing (HPC) machine with 128 GB of RAM and 4 quad-core Intel(R) Xeon(R) CPUs, E7420 2.13 GHz. For each giga-pixel tile at its input and with data quantised to 8 bits/pixel the system required:

- 38 sec. for building the two hierarchical image representation structures;
- 4 sec. for the structure polychotomy into two sets of 256 attribute-zones each;
- 16 sec. for the CSL model computation;
- 1 sec. for the output image generation.

This sums up to approximately 1 min./GB of data at 0.5m. spatial resolution. The average memory requirement for each tile was approximately 30 times the input size, i.e. 30 GB of RAM. The involved algorithms run concurrently and the code pluralisation is a product of the High Performance Image Processing Algorithms for Remote Sensing (HIPARS) collaboration between the JRC and Rijksuniversiteit Groningen, Netherlands.

Chapter 4
Experimental setup

4.1 Input image data available

The IQ system is designed to be operational, as described in the introduction. The satellite and airborne data used were acquired with optical sensors with a spatial resolution of 10m or smaller in order to allow detection of single buildings or groups of buildings. The data are hosted in the Community Image Data Portal[1] (CID). The CID Portal is a web portal to search and access remote sensing data and derived products hosted at JRC for a variety of applications. In this study, we use in total 11438 panchromatic and multi-spectral satellite data sets from SPOT-2 and SPOT-5, RapidEye, CBERS-2B, QuickBird-2, GeoEye-1, WorldView-1 and WorldView-2. In addition, airborne data sets covering for example entire Guatemala were available as RGB bands. The number of data sets per sensor and the area covered is detailed in Table 1.1. The different data sets cover a wide range of spatial resolutions from 15 cm airborne data sets to 10m of the SPOT-2 sensor. Radiometrically the entire visible and near infrared part of the spectrum is covered with wide panchromatic bands and up to eight multispectral bands of WorldView-2. Fig. 4.1 depicts the spatial/spectral distribution of sensors used in this study. In addition, some data sets consist of pan-sharpened multispectral images with the spatial resolution of the panchromatic band. It is important to note that this often changes the radiometry of the data set, which has to be taken into account during further processing.

4.2 Ancillary information

Several additional data sets were used in the workflow as ancillary data in the processing or for the validation of the processing output. For the orthorectification of some of the satellite data we used the TerraColor[2] as reference layer. This is an orthorectified global imagery base map at 15m spatial resolution built primarily from Landsat 7 satellite imagery. The Open Street Map[3] (OSM) data were used to extract a high resolution land-sea border. During the processing two global data sets are

[1] URL: `http://cidportal.jrc.ec.europa.eu`

[2] URL: `http://www.terracolor.net`

[3] URL: `http://www.openstreetmap.org/`

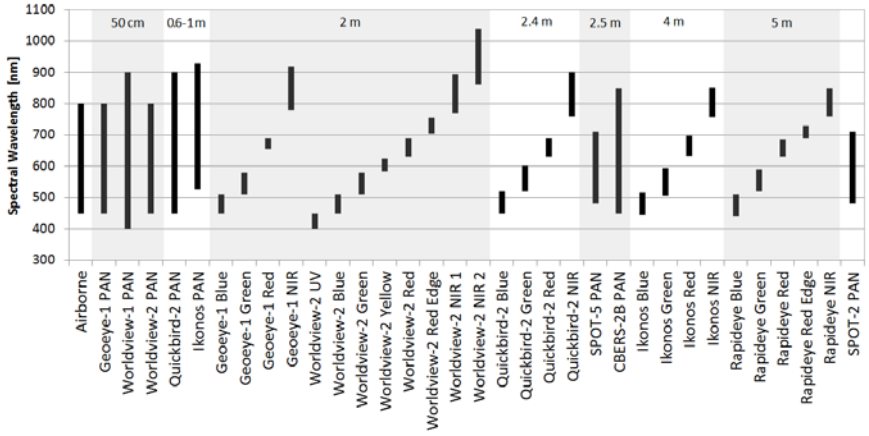


Fig. 4.1 Spectral coverage of sensors used in the study. The satellites cover a wide spectral range in the visible and NIR part of the spectrum. The spatial coverage includes various resolutions form 15 cm airborne to 10m panchromatic images of SPOT-2.

used to control the information extraction. The first one is the urban class of the MODIS Land Cover Type product [SFP10]. The second is the LandScan (2008 & 2010)[4] high resolution global population data sets were used.

4.3 General workflow

The main elements of the I2Q workflow characterising the GHSL experiment discussed here are four: the input image data, the reference data set, a pre-processing module, and a processing module (Fig. 4.2). The input image represent the data to be processed. The reference set has the crucial function to support the optimisation of the spatial and thematic consistency during the GHSL production. Currently, it consist of i) a global mosaic of Landsat image data ii) OSM vector data, iii) global LandScan gridded population data (LSPOP) of 2010 at a resolution of circa 1km, iv) global MODIS urban areas at resolution of circa 500meters, and v) a Built-Up Reference (BUREF) data layer containing the best estimation of presence of built-up areas at the GHSL global scale. All these reference information layers are available for the pre-processing and the processing I2Q modules trough WMS and TMS standard protocols.

Before the first run of the I2Q GHSL workflow, the BUREF layer was calculated by merging the most consistent information available globally: the LSPOP and MODIS information layers. Independently from that initial choice, the whole system is designed having an incrementally evolutionary approach: the output of any given image information extraction run/experiment, if passing validation and consistency check, will contribute to improve (thematic accuracy, spatial/temporal completeness) the available BUREF layer. The expectation is that this retro-action

[4] copyright by UT-Battelle, LLC, operator of Oak Ridge National Laboratory under Contract No. DE-AC05-00OR22725 with the United States Department of Energy

mechanism will contribute in enhancing stepwise the overall reliability and completeness of the GHSL output.

The pre-processing module basically performs two functions: i) checking and optimisation of the spatial consistency of the input image data and ii) checking and flagging of potential occlusions and no-data areas in the images. The spatial consistency is optimised by using an available reference set with 15-m-spatial resolution and an expected RMS spatial tolerance of around 20 meters, while the occlusions and no-data areas are detected by an internal recognition mechanism. Details on these steps are available in Section 5. Because of the characteristics of the spatial reference layer available, the current GHSL production is not fitting with the 1:10K local scale specifications, and consequently only the 1:50K regional and 1:500K global will be delivered for the moment.

Note that no spectral calibration steps are implemented in the current workflow. Thus all the subsequent processing steps work with spectrally un-calibrated satellite image data input. This choice was dictated by the fact that an important part of the input image data volume was not suitable for being radiometrically calibrated with existing tools, requiring the availability of parameters for each scene to derive top-of-atmosphere reflectance. Apart from some VHR optical sensors, in particular this information was unavailable for pan-sharpened multispectral SPOT-5 image data and for airborne/satellite mosaics that were instead playing an important role in the current experiment, both as volume of data involved and interest of users in the output.

The absence of radiometric calibration did not allow including spectral image descriptors derived from multi-spectral band combinations such as the Normalised Difference Vegetation Index (NDVI), that may significantly increase the BU/NBU discrimination in case of availability of multispectral image data. New spectral calibration methods able to process a more complete range of input data are under study and will be applied in next I2Q GHSL workflow releases. Nevertheless, the capacity to extract image information without using spectral band combinations can be seen as a benefit for the applicability of the proposed workflow, allowing the processing of mono-channel image data.

4.4 Image information query infrastructure

The I2Q System architecture relies on different components interacting between each other to allow massive image elaboration and validation of the results (Fig. 4.2). The I2Q System main modules are:

1. Web server;
2. Image server;
3. Application server;
4. Database server;
5. Elaboration core (computer cluster + image query scripts + image tiling + image statistics)

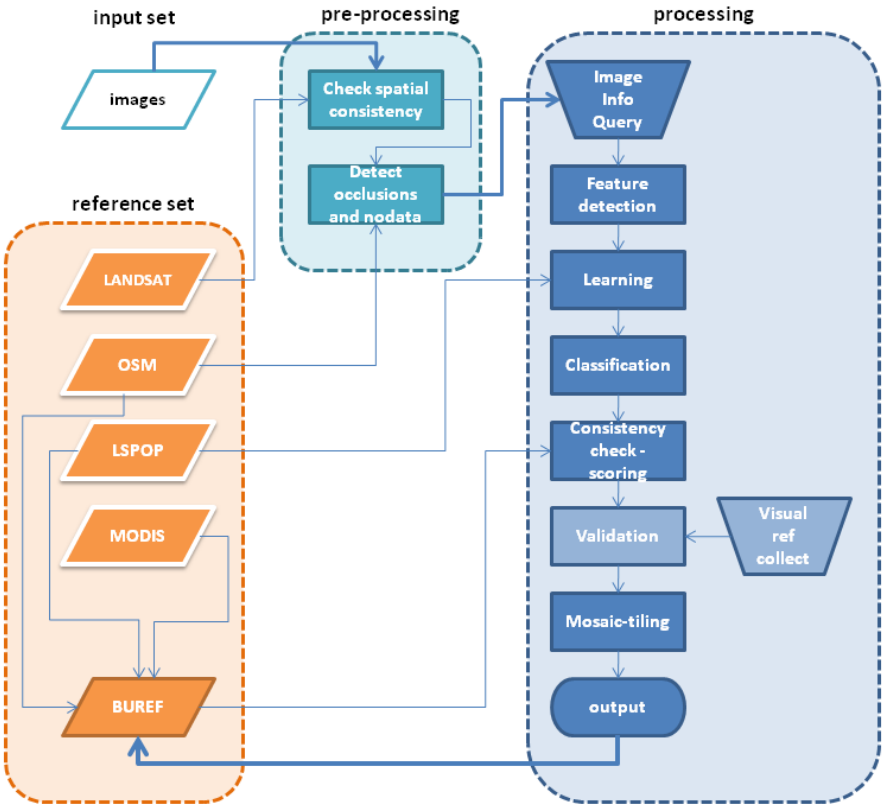


Fig. 4.2 The general I2Q GHSL processing workflow.

4.4.1 Web server

This component provides the access to the Graphical User Interface of the system using a web browser. Through this interface, the user can interact with the system and perform different operations like:

- Submission of image processing requests and monitoring of the elaboration status;
- Visual inspection of the elaboration output and comparison with the reference image or any other correspondent image/information layer present in the system;
- Implementation of the validation protocol;
- Query different image information layers available through WMS/TMS protocols.

The user interface is composed of the following main areas:

- **IQJobs:** this is the area dedicated to the image processing. Users can submit jobs and monitor their status;
- **IQMaps:** this section is dedicated to the visual inspection of the elaboration output. Here it is possible to display the processed scenes and compare them with the reference image or any other correspondent image/information layer

present in the system. IQ Maps is also used for the visual interpretation during the collection of reference data during the validation process;
- **IQValidation:** this area is dedicated to the validation protocol. Here it is possible to define which scenes are part of the validation process and the number of statistical samples;
- **IQServices:** this part is dedicated to the image information layers provided through WMS/TMS protocols. From the services area it is possible to query different types of information layers for computational purposes or use them as a background for visual comparison.

4.4.2 Image server

This component optimises image delivery through the web server. Moreover it provides the possibility to perform advanced operations on the images like: i) Image adjustment, reprojection and query, ii) Management and navigation of very large image sets, iii) Advanced cartographic output and support of Open Geospatial Consortium (OGC) standards.

4.4.3 Application server

This element provides an additional software layer between user requests and the database server. This provides the following advantages: i) Centralised module for implementing business logic of the system, ii) Increased security and performance in data access and management, and iii) Optimisations in application development.

4.4.4 Database server

This component stores all the information needed for the functioning of the system plus the information related to the datasets available, and in particular: i) Processing information and status, ii) Statistical data and validation results, Area of Interest (AOI) of the images and metadata.

4.4.5 Elaboration core

This part is in charge of executing the different types of processing required by the users and those of the system. To fully exploit the computing power and the resources available, a high-performance computing cluster has been implemented. In line with the hardware also the software was implemented allowing parallel processing (see details in Sec. 4.4.6. An important task of the elaboration core is the image tiling (image subdivision by a regular grid). It is necessary in order to render the images at different zoom levels without affecting the performance of the system. The tiling is made by a specific script that is able to store the so called tiles directly

on the file system or in a database. There are different technologies and philosophies behind tiles generation and provisioning, the one adopted is to pre-generate the tiles up to a certain zoom level, while the tiles generation for the next levels is done on the fly.

4.4.6 Infrastructure

The IQ system infrastructure is heterogeneous and scalable. The project started with a small number of blade servers and limited storage capacity that progressively increased together with its processing power. Currently the I2Q system infrastructure is composed as follows: nine blade servers (two of which are host for virtual machines); two workstations equipped with GPU; five virtual machines; two storage systems for a total of 3 TB net capacity, for a total of 112 CPU cores and 296 GB of RAM.

Chapter 5
Preprocessing

5.1 Positional accuracy

The positional accuracy of each input scene and the positional consistency between any pair of overlapping scenes are fundamental quality measurements needed during the generation of the GHSL layer and for any subsequent use. This is a challenging issue given the heterogeneity of the input data sets and the absence of global reference layers matching the highest resolution of the input imagery (i.e., 50cm). The best optical imagery reference layer with reported accuracy across the globe and freely available is the Landsat 7 nearly global coverage [TGD04]. This freely available imagery needs to be mosaiced to ease accuracy measurements of arbitrary HR/VHR images across the globe. Although we have developed a method for mosaicing large data sets [Soi06; BS05] and have applied it at continental scale for Europe with Landsat data [BGS08] and IRS LISSIII plus SPOT-4/SPOT-5 imagery [Soi08], at global level we currently use TerraColor as our imagery reference layer.

5.1.1 Positional accuracy against TerraColor

Except for CBERS-2 HRC imagery (see next paragraph), we have assumed that the geolocation of the input imagery as indicated by the image metadata is correct. The positional accuracy is determined using normalised cross-correlation measurements [BS72]. That is, given a square template cropped from the input image at a given position, normalised cross-correlations are calculated between this template and equivalent templates cropped in the reference image with the centre pixel falling within a search window centred on the same position as that of the original template. The vector separating this position from the position at which the maximum value of the cross-correlation occurs is used as a local estimation of the translation vector between the input and reference images at that position. If the normalised cross-correlation is not circular enough or not high enough, the estimated translation vector is deemed unreliable and discarded. The mean, RMSE, and standard deviation of the horizontal and vertical displacements are then reported. Subpixel measurements are obtained thanks to a quadratic interpolation of the correlation function [TH86]. Because the spatial resolution of TerraColor is 15m, only errors above 1.5m (the

theoretical smallest displacement that can be measured with subpixel measurement using quadratic interpolation is a 10th of the resolution of the reference) are significant. This is enough for assessing the quality of the final GHSL layers of regional and global scale, which are delivered at a maximum scale of 1/50,000.

For the special case of CBERS-2 HRC imagery, the geolocation of the input imagery is unreliable with scene centre displaced by up to 40 kilometres [dCE04; Cas+11]. In this case, we have used the procedure described above to automatically find tie points necessary to correct the geolocation of the input CBERS imagery. This is achieved by subsampling the CBERS images at the resolution of TerraColor. Displacement vectors are then extracted by considering only those vectors as reliable that clearly cluster in well defined regions of the x-y space. The CBERS image is then warped using the affine transformation model with least square estimation of the transformation parameters given the final set of tie points. This procedure is summarised in Fig. 5.1.

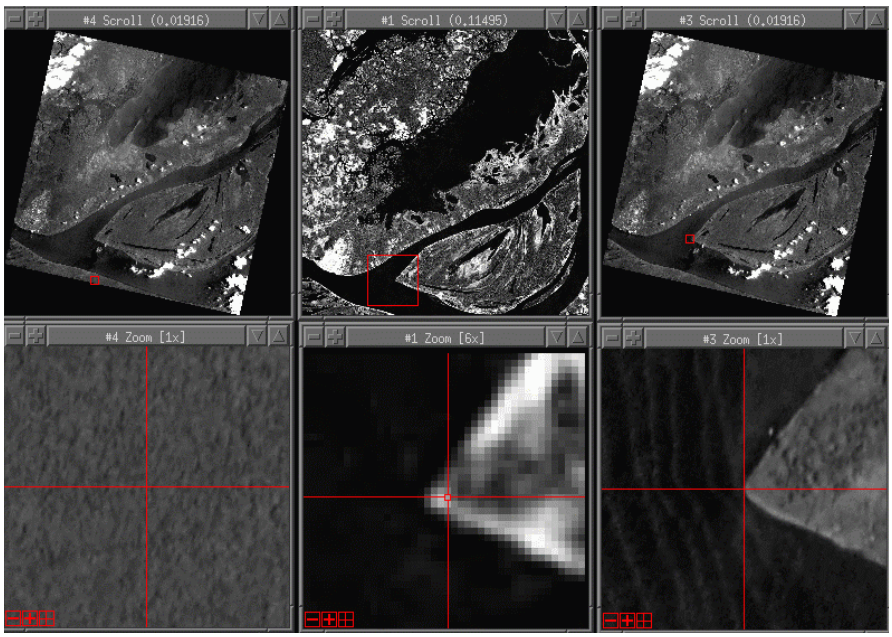


Fig. 5.1 CBERS-2 HRC geolocation using TerraColor as reference. The red cross on the zoom images (bottom image row) indicates the same geolocation in all three images. Left column: original CBERS scene. Middle column: TerraColor in the same domain as that specified by the original CBERS scene (red channel). Right column: the CBERS scene warped to the TerraColor image.

5.1.2 Relative positional accuracy (consistency)

The relative positional accuracy (i.e. consistency) between any pair of overlapping images is calculated similarly to the positional accuracy. It is a consistency measurement in the sense no reference is used but merely a comparison between two data sets in their overlapping domain.

5.2 Cloud detection

Precise cloud detection remains a difficult topic per se due to the absence of formal definition of a cloud that can be translated into an image processing chain with well defined parameter values. For instance, 'What is the threshold level to declare whether there is a thin cloud or not?' or 'Where is the precise location of the boundary of a cumulus?' are questions that will probably never be answered. Nevertheless, cloud masks indicating the image areas most affected by clouds can be generated by a wide variety of methods. For example, cloud detection in medium resolution imagery with sensors offering a wide range of spectral bands including thermal bands is usually based on the Automatic Cloud Cover Assessment (ACCA) method [Iri+06]. This type of method requires not only a wide range of spectral bands but also precise Top-Of-Atmosphere (TOA) reflectance values. The parameters necessary for TOA reflectance calculations, such as sensor gain and offset parameters and solar irradiance at the given wavelengths, are not always precisely known for HR/VHR satellite data. Experiments detailed in [Soi08, Chap. 3] have shown that ACCA can be adapted for sensors with green, red, near infrared, and short-wave infrared bands such as SPOT-4 HRVIR, SPOT-5 HRG, and IRS-LISS III sensors. However, for VHR data that sometimes contain only 1 panchromatic channel, this path is not viable. Therefore, an ad hoc method for generating cloud masks has been developed. More precisely, rather than relying on calibrated data, we have developed a method that translates a visual definition of a cloud into an image processing chain. There are actually two chains, the first for situations where only one VHR band is available, the other for all other cases (i.e., multi-band HR/VHR data).

5.2.1 Panchromatic VHR data

This technique was originally developed for CBERS-2 HRC panchromatic images. It exploits the fact that the cloud boundaries consist of a smooth transition from bright to darker regions spanning over many pixels. Smooth transitions are detected using point-wise arithmetic differences between thick morphological gradients [RSB93] of increasing size. The non-zero values of the resulting image delivers a mask of the cloud boundaries. The holes of this mask are than filled by a morphological reconstruction by erosion called the fillhole transformation [Soi03]. Finally, a dilation by a disk of fixed radius (15m) is applied to make sure that most pixels corrupted by the detected clouds are indeed covered by the cloud mask. Figure 5.2 shows a CBERS-2 HRC image and its corresponding cloud mask.

5.2.2 Multispectral HR/VHR data

For multispectral HR/VHR data (typically four channels), the adaptation of the ACCA method proposed in [Soi08, Chap. 3] did not provide satisfactory results in many cases due to the unavailability of precise calibration parameters. For this reason, we have developed a method taking into account the multispectral nature of the input data relying on a visual characterisation translated into an image process-

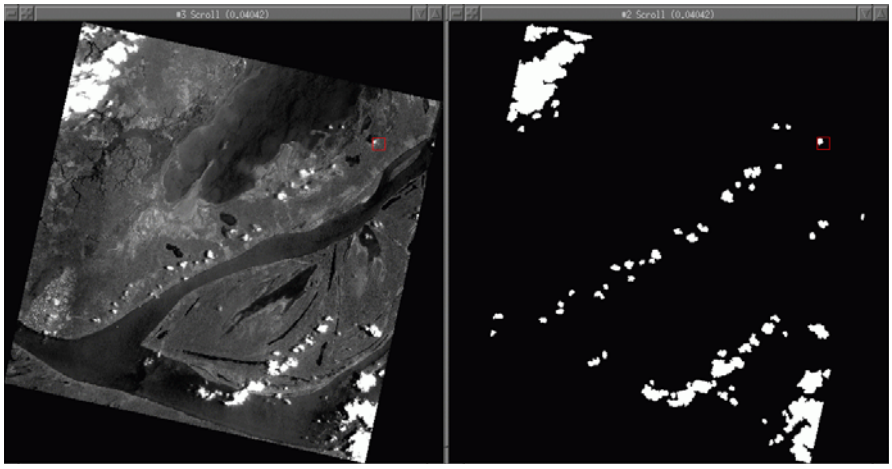


Fig. 5.2 Cloud detection from panchromatic data: example on a 2.5m CBERS-2 HRC scene over Brazil.

ing chain. The method relies on fact that clouds appear has regional maxima [Soi03] in all visible and infra-red channels of multispectral images. Therefore, the pointwise intersection of the extended regional maxima obtained for each channel is used at the basis of an image processing chain delivering a mask of potential clouds.

5.3 Land mask

A global land mask was generated using OpenStreetMap[1] (OSM) data. The coastline data was downloaded from OSM services and converted into a Boolean layer. It was then inserted into PostgreSQL-PostGIS database. Thanks to the image server (MapServer), which provides all OGC standards with all authorised output formats, it is possible to query the land mask layer through standard WMS with the desired parameters such as those indicating the image extent, resolution, projection, and output format. The resulting mask was used, for instance, during the positional accuracy measurement to avoid calculating correlations in the sea or during learning on textural features to discard regions covered by the sea.

[1] URL: `http://www.openstreetmap.org`

Chapter 6
Feature extraction

This chapter gives an overview of the different types of features collected from the available imagery as a basis for a medium level semantic layer describing the human presence through the evidence of built-up.

6.1 Textural features

The textural image features used in this study are derived from grey level co-occurrence matrix (GLCM) contrast textural measurements [RKI73]. The GLCM matrix is a $n \times n$ matrix containing the relative frequencies with which two pixels linked by a spatial relation (displacement vector) occur on a local domain of the image, one with grey level i and the other with grey level j, with $i, j \in [0, \ldots, n-1]$, where n is the number of grey levels with which the image has been coded. The contrast textural measurement is formalised as follow:

$$CON = \sum_{i=1}^{n} \sum_{j=1}^{n} (i-j)^2 \cdot P_{i,j},$$

where n is the number of grey levels present in the image, and $P_{i,j}$ is the $(i,j)th$ entry of the co-occurrence matrix. The contrast textural measures calculated using anisotropic displacement vectors are combined in a rotation-invariant image feature called PanTex [PGK08] by using extrema operators. In [Pes+11], it was demonstrated that PanTex is strongly correlated with the presence of buildings as well as with their spatial density [PHO11]. The capacity of PanTex to discriminate BU/NBU areas is mainly linked to the fact that it is a corner detector [GSP12]. In addition, the built-up areas generate high spectral heterogeneity in the local domain due to the heterogeneity of materials used and because of the fact that buildings are generally casting shadows. In this study, PanTex method was improved by adding i) an a priori weighting mechanism substituting the usual standardisation step before the integration of different displacement vectors and ii) the parametrisation of the radius used for generating the displacement vector list.

The basic parameters necessary to calculate the PanTex feature are the window size $wsize$ and the list of displacement vectors d used for generating the GLCM. $wsize$ is analytically derived from the GHSL scale of the expected output that in

this experiment is set at the 1:50K GHSL technical specifications. Consequently, $wsize = 50$ meters that is corresponding to the spatial unit of the GHSL regional scale. The $wsize$ will be translated in corresponding number of image pixels by ratio with the image spatial resolution information. In this study, the displacement vectors are generated assuming a radius of 10 meters that corresponds with the maximum size of the image pixel tested during the experiment, and fits well with a priori knowledge regarding the expected size of the majority of the built-up structures. Table 6.1 summarises the $wsize$ and the number of displacement vectors d for the most common image resolutions used in the study.

Table 6.1 PanTex parameters used in the study.

Image resolution (m)	Number of displacement vectors	Window size (pixels)
10.0m	4	5
5.0m	6	10
2.5m	16	20
1.0m	28	50
0.5m	56	100

6.2 Morphological features

Radiometric and textural features describe material properties and object patterns, respectively. They are used as indicators of human presence. For each given scene the features are employed as markers to validate building footprint candidates. The latter are summarised in a separate information layer that is a product of a multi-scale morphological analysis protocol referred to as the "mtDAP" (max-tree Differential Attribute Profile).

The mtDAP protocol [OPS12] computes the Differential Attribute Profile (DAP) vector field [Dal+10] from the input imagery. DAPs are non linear spatial signatures that are used extensively in remote sensing optical image analysis in ways analogous to spectral signatures. The DAP of a pixel is the concatenation of two response vectors. The first registers intensity differences, i.e. contrast, within a top-hat scale-scape of an anti-extensive attribute filter γ, and the second intensity differences on the bottom-hat scale space of an extensive attribute filter ϕ. The pair (γ, ϕ) defines an adjunction with γ typically being a connected attribute opening and ϕ being a connected attribute closing.

The mtDAP can be configured with any morphological attribute filter but in this case, simple area openings and closings prove to be sufficient. The area attribute is used to order objects based on size and it is computed incrementally. The protocol consists of three core modules; the hierarchical image representation using a Max-Tree and a Min-Tree structure [SOG98], the attribute zone tree-polychotomy scheme and the spatial signature export module. A brief description of each module follows.

Hierarchical image representation schemes [SN12] aim at organising the image information content into meaningful structures or components and registering component transitions through their nesting properties. Examples in morphological im-

age analysis are the Max-Tree [SOG98], the Component-Tree [Jon97; Jon99], and the Alpha-Tree [OS11; OS12]. This work makes use of the Max-Tree structure for computing anti-extensive attribute filters on grey-scale images. The Max-Tree is a rooted and uni-directed tree in which every node N corresponds to a single *peak component* that associates to a set of *flat zones* [SS95]. A peak component at level h is a connected component C, i.e. a connected set of maximal extent, of the binary set given by thresholding the input image at level h. A flat zone is an image region consisting of iso-tone and path-wise (or otherwise) connected image elements. If the full extent of a peak component coincides with a single flat zone, the component is referred to as a *regional maximum*. The leaves of a Max-Tree correspond to the regional maxima of the input image and every node points to its parent P that corresponds to the first superset C_P of C_N at level $h_P < h_N$. The root node corresponds to the image background and points to itself.

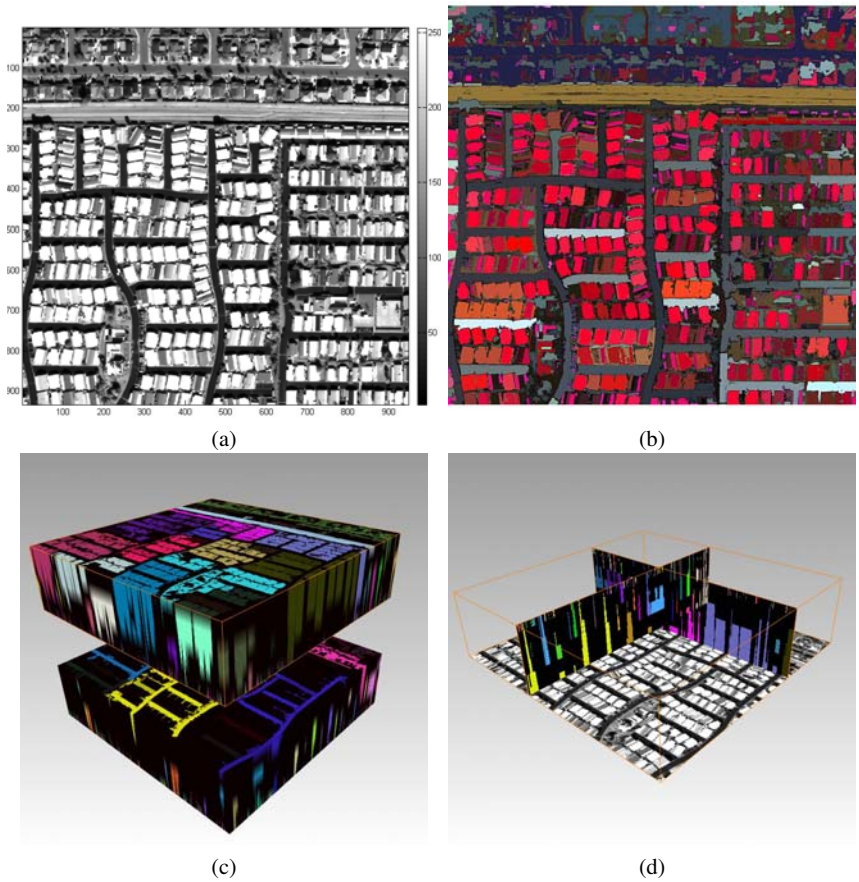


Fig. 6.1 Example of DAP vector fields: (a) the input image; (b) the colour representation of the DAP vector field using the CSL model; (c) the DAP vector field in colour-map projection in which the two volumes correspond to the opening top-hat and closing bottom-hat scale space respectively; (d) a cross section of the two.

The Max-Tree treats bright structures as foreground information with respect to a darker background. The inverse is represented by a Min-Tree structure, i.e.

foreground information are dark components resting against a brighter background. The Min-Tree is equivalent to a Max-Tree on the inverted input image.

Computing both structures on the input image offers the possibility of evaluating the significance or extent of the contribution of each connected component to the structured image information content with respect to the nesting order. This the key idea behind the DAP decomposition, which assumes size as the dominant criterion for ordering components. To obtain the DAP vector field, i.e. the set of all DAPs for a given image, the mtDAP protocol uses the Max-Tree to compute the top-hat scale-space of an area opening and the Min-Tree to compute the bottom-hat scale-space of an area closing. Instead of re-iterating the same operator configured with a varying attribute threshold for each scale space, the mtDAP employs the "one-pass" method [OPS12], which generates an area zone polychotomy of each respective tree; each node is assigned a unique area zone identifier in a single pass through each structure.

The DAP vector field can be visualised as a 3D set, in which every $x - y$ plane, corresponding to a particular scale of the decomposition, stores in each pixel position the pixel's contrast computed in the respective top-hat or bottom-hat. An example is shown in Fig. 6.1. The grey level image (a) is represented in a colour coded CSL model showing the building footprint candidates (b, see next section for details). Fig. 6.1 (c) and (d) show three dimensional visualisations of the DAP vector field, in which the top volume set in Fig. 6.1 (c) corresponds to the top-hat scale-space and the bottom volume set to the bottom-hat scale-space. Exporting the DAP vector field requires a single pass through the image definition domain during which, for each pixel visited, its node ID is retrieved. Visiting the respective node gives access the area zone ID and contrast with respect to the zone's highest attribute boundary. The type of the tree is used to decide on which of the two volume sets the pixel is be reported, the area zone ID to decide on which plane it belongs to and the contrast to assign a value on the corresponding coordinates on that plane. A cross-section through the two volume sets is shown in Fig. 6.1(c), where the scale-based responses of a sample set of pixels are shown.

6.3 Feature compression and storage

The (scale) resolution of DAPs, i.e. λ vector length and the between-scale spacing, is a critical parameter in their utilisation as feature descriptors. It is typically set either empirically or based on the outcome of some statistical learning procedure. Evidently, higher spatial input resolution offers a more detailed profile for each pixel under study. A drawback in this case is that by increasing the vector length I the number of DAP vector field planes (Fig. 6.1(c)) increases proportionally, i.e. $2 \times (I - 1)$. When dealing with large data-sets this can be prohibitive. To counter the resulting data explosion a compression model was devised to radically reduce the dimensionality of the DAP descriptors. It is called the *Characteristic-Saliency-Level* or *CSL Model* [POG12] and is a medium abstraction semantic layer that can be projected on the HSV colour space for the visual exploration of built-up extracted from VHR satellite imagery, see Fig. 6.1(b).

The CSL model is a non-linear mixture model consisting of three characteristic parameters extracted from the DAP of each pixel. The Characteristic (C) is the min-

imum scale at which the maximal contrast response is observed in the DAP. The contrast value is the Saliency (S) and the level (L) is the highest peak component level from which this value is computed. The model is computed directly from the polychotomy of the two tree-based data structures and with no need for exporting the DAP vector fields. It reduces radically the dimensionality of the DAP vector field to a three-band representation in a statistical-model free approach, i.e. it avoids clustering based on the statistical distribution of the DAP features of a given image. It does not require manual tuning and its computation is independent of the length of the DAP. This makes it suitable for automated, user-independent applications like the GHSL.

The colour-mapped CSL model provides a pool of building footprint candidates like in Fig. 6.1(b). Warmer colours indicate higher confidence that a particular structure is a building. The selection of candidates employs the thresholded PanTex built-up indices as markers in order to draw the final high-level semantic layer containing all targeted built-up.

Both the compressed multi-scale morphological (CSL) and textural (PanTex) image descriptors are aggregated and stored with the spatial resolution corresponding to the finest scale supported by the specific image information query, that in this case is equal to a grid of 10 meters, corresponding to the GHSL 10K specs. They are then ready to feed the subsequent image information queries and classification with the maximum of spatial detail. With this schema, the heterogeneous set of input sensor/platforms with different resolutions generates raw image features also heterogeneous in spatial resolution but then they are spatially standardised to the 10-m nominal grid. In many cases also this operation of spatial aggregation will introduce a compression of the data space with positive effects in the overall performances of the system (storage, indexing, and retrieval). For example, an input image feature generated with VHR image of 0.5-m-resolution, if aggregated to 10 meters will show a compression factor of 1:400.

Chapter 7
Learning and classification

7.1 Learning approach

We can distinguish three different modes of learning and classification implemented in the current I2Q system: i) adaptive learning, ii) meta-learning, and iii) discovery. The different learning approach are used in order to mitigate the drawbacks of admitting low-quality (low resolution, low accuracy) reference data driving the learning phase. There may be situations, where the learning form a too coarse or outdated reference may lead to an omission of settlements that were detected correctly. This capacity is crucial when processing multiple-scene image input data without manual intervention in the input parameter set.

In the 'adaptive learning' mode, the system optimises the decision thresholds of the input features using a given reference layer. The 'meta-learning' mode is used to study the behaviour of these decision thresholds in a set of scene processed and to detect regularities: for example typical thresholds for a given sensor in specific regions. The output of the meta-learning is then exploited during the 'discovery' mode that can be activated in order recover image information lost due to errors (incompleteness, inconsistencies) in the reference data, or due to different scales of generalisation between the image-derived information and the reference data. In practice, the adaptive learning optimises consistencies between the image information and the reference data. The meta-learning and discovery modes allow detection built-up in cases where reference data is not available with the necessary thematic and/or spatial precision.

The typical workflow combining the three modalities will be as follows: i) run the adaptive learning in all available scenes and classify the extracted features based on the available reference data, ii) run meta-learning in the set of successfully classified scenes with available reference data, and iii) run discovery mode in the set of scenes under a given threshold of quality detected after the adaptive learning phase.

7.2 Adaptive learning on textural image features

The PanTex method [PGK08] (see also Sec. 6.1) provides measurements correlated with built-up density, which however depends strongly on the specific image contrast and the sharpness of the input imagery. While the measure has a high dis-

Fig. 7.1 Example of input image data (CBERS 2.5m), Brazil.

crimination power of BU/NBU area in a variety of conditions [PE09], the optimal classification threshold is strongly dependent of the image characteristics and requests manual interaction. In an attempt to automate this procedure, [PE09] tested a normalisation of the image grey level distribution with a unique global classification threshold, but the results were suboptimal.

As opposed to this signal-driven approach, we propose here the use of third party sources of information to derive the optimal classification threshold for each image. Since we are aiming at a fully automated process for a global coverage, also the reference data must be global, complete and consistent to support the adaptive learning. The LandScan population density layer [Dob+00] and the MODIS500 [SFP10] 'Urban Areas' layer were identified as a source of information fulfilling the previous constraints and with a strong correlation with the BU areas. While LandScan layer represents the density of people in cells with a spatial resolution of approximately 1km (30 arc second$\times$ 30 arc second), the MODIS500 layer represents the presence of 'Urban Areas' estimated by classification of MODIS input data, with a resolution of approximately 500x500 meters.

Accordingly, a method exploiting the 1km-resolution LandScan and 500m-resolution MODIS500 layers for deriving the optimal classification threshold from the 10m-resolution PanTex measurements is described.

7.2.1 PanTex Learning on LandScan

This section describes the learning of the best image-derived PanTex thresholds for the discrimination of BU/NBU areas from the LandScan population density reference layer.

Let PanTex feature dataset be denoted by x and the raster layer with the LandScan reference in the bounding box of x be denoted by y. Both datasets have different resolutions, such that one pixel q of y encompasses several pixels $q^x = \{r_i\}$ of x. This is illustrated in Figs. 7.2 and 7.3. The PanTex features were extracted from the input image shown in Fig. 7.1. While the spatial correlation is obvious, the links

between the PanTex and the LandScan layers must be modelled carefully. In order to cope with the uncertainty of the information in y which is fused into x, a statistical model is employed.

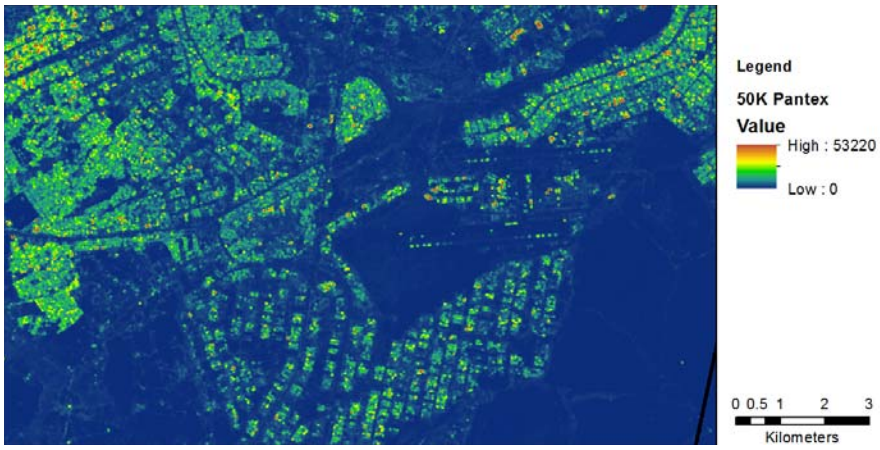


Fig. 7.2 Unscaled PanTex features derived from the input image in Fig. 7.1.

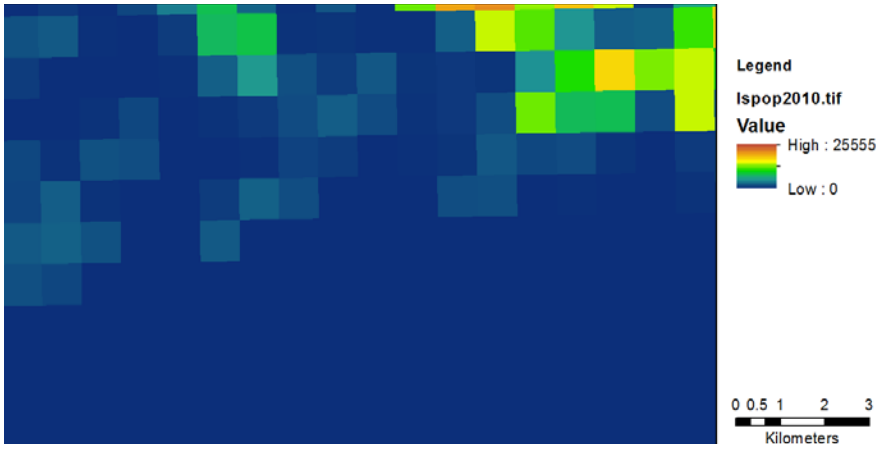


Fig. 7.3 LandScan 2010 population layer corresponding to the image area in Fig. 7.1. This layer is used for the adaptive training. Note the differences in the spatial resolution.

It can be assumed that the BU spatial density increases with the density of population until the BU density reaches a saturation s (i.e. an area is entirely built-up). Additional increase in population density can only be accommodated in vertical building growth, which however does not change the BU density anymore. This observation led to a model between the BU density π_{BU} bounded by s and the density of population in a cell q:

$$\pi_{BU}(y(q) \mid \lambda, s) = s(1 - e^{-y(q)/\lambda}),$$

where λ determines the saturation point. This break point λ depends on the geographic location (including cultural/economic setting), but it can be inferred from the observations x and y.

By analysing the formation of the PanTex measurement, a statistical model of the realisations of x is selected. Assuming that the grey levels of the input image follow an identical Gaussian distribution, the differences (contrast) between pixels values follow also a Gaussian distribution. The contrast textural measurement is a texture dependent linear combination of the squared differences, such that it follows a χ^2-distribution or more generally a Gamma distribution. Finally, the PanTex is a non-linear merge of those measurements, but does not allow a close distribution form. Experimentally, the PanTex values are well approximated by a Gamma distribution in pure BU or pure NBU areas. Thus, the distribution of the PanTex values can be modelled as a mixture of Gamma distributions in an unknown environment:

$p(x(r)) = P_{BU} p_\Gamma(x(r) \mid \theta_{BU}) + P_{NBU} p_\Gamma(x(r) \mid \theta_{NBU})$,

where $P_{BU} + P_{NBU} = 1$ are the prior probabilities of BU and NBU, $x(r)$ is the PanTex measure at location r, and $\theta_{BU}, \theta_{NBU}$ are the Gamma distribution parameters in the BU and NBU areas. The prior P_{BU} can be assimilated as a density of BU, and it is approximated by π_{BU}. By inserting the information of the density of population, the probability distribution of the pixels $q^x = \{r_i\}$ falling into q is structured as follows:

$p(x(r_i) \mid y(q), \lambda, s, \theta^q_{BU}, \theta^q_{NBU}) = \pi_{BU}(y(q) \mid \lambda, s) p_\Gamma(x(r_i) \mid \theta^q_{BU}) + (1 - \pi_{BU}(y(q) \mid \lambda, s)) p_\Gamma(x(r) \mid \theta^q_{NBU})$,

where the Gamma distribution parameters depend on the location q. In summary, the joint probability of both observations x and y is structured and parametrised in the following way:

$$p(x, y \mid \lambda, s, \{\theta^q_{BU}, \theta^q_{NBU}\}) = \prod_q p(y(q)) \prod_{r \in q^x} p(x(r) \mid y(q), \lambda, s, \theta^q_{BU}, \theta^q_{NBU}), \quad (7.1)$$

where $p(y(q))$ is the distribution of the density population values, which can be approximated by a histogram. The parameters $\lambda, s, \{\theta^q_{BU}\}, \{\theta^q_{NBU}\}$ of this model can be derived by expectation-maximisation [Moo96]. No spatial constraints are imposed on this model, while the data seem to be spatially correlated. To gain in robustness, the multi-dimensional image formed of the parameters $\theta^q_{BU}, \theta^q_{NBU}$ is low-pass filtered at each iteration of the Expectation-Maximisation. This process benefits from the context, giving higher confidence in the estimates. While the parameters model the link between both data x and y, it also embeds crucial information for classifying the pixels of x into BU or NBU. In particular for a pixel r belonging to a cell q, its classification is given by:

$\arg\max_{A \in \{BU, NBU\}} \pi_A(y(q) \mid \lambda, s) p_\Gamma(x(r) \mid \theta^q_A)$.

Finally, this method allows using additional source of information, the density of population, in order to derive location adaptive decision for detecting BU in unknown environments. As the density of BU is not derived from the observation x only, the algorithm does not fall in unwanted local maxima.

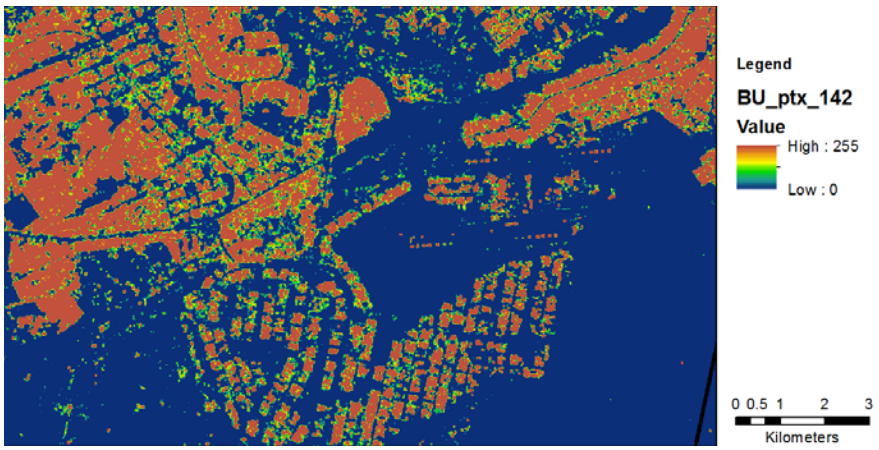


Fig. 7.4 PanTex features of Fig. 7.2 rescaled with the LandScan2010 data (Fig. 7.3) as reference.

7.2.2 PanTex Learning on MODIS

Let the data of 'Urban Areas' be denoted by y and the PanTex feature existing in the bounding box of y be denoted by x. Both datasets have different resolutions, such that one pixel q of y encompasses several pixels $q^x = \{r_i\}$ of x. Let $\bar{x}_{BU}$ and $\bar{x}_{NBU}$ with $\bar{x}_{BU} \geq \bar{x}_{NBU}$ be the average of PanTex feature in spatial domain defined by the urban areas and their complement, respectively, of the same scene under processing, being the urban areas $y_{BU} \neq \emptyset$. Then the rescaling of the PanTex measurements to the BU-NBU decision map is done by linear rescaling of x such that $x' = (x - \bar{x}_{NBU}) / (\bar{x}_{BU} - \bar{x}_{NBU})$ with $x' \in |0, 1|$.

In addition to the learning with the LandScan data, we have implemented two learning options using the MODIS 'urban areas' as reference (Figures 7.5). The MODIS 500m map of global urban extent was generated by automatic classification of MODIS multispectral data using regression tree techniques and dedicated models fitting with specific characteristics of the different ecological regions [SFP10]. The MODIS urban layer is the most detailed and the most consistent global information layer available today being produced from the same sensor and the same methodology. Moreover, some work on validation of this source was also reported in [SFP10]. Nevertheless, two main drawbacks were expected in using this source for training of the HR, VHR image data used in this study: i) the resolution of the sensor and the techniques used for image information extraction would presumably underestimate the detection of scattered and/or vegetated settlement patterns, introducing a bias in favour of non-vegetated compact and large settlements, ii) in some areas there is a shift in the MODIS data with respect to the HR, VHR data available, which can be larger than 2-3 times the MODIS pixel size.

The two options of PanTex learning from MODIS are 'by area matching' and 'by ROC (Receiver Operating Characteristics) optimisation', respectively. The option 'by area matching' (Fig. 7.6) minimises the difference between the total surface of the BU areas estimated in the reference and in the pantex-derived layers, by exploring a range of thresholds in the PanTex image features. The option 'by ROC optimisation' (Fig. 7.7) instead collects the PanTex threshold that minimises the

Equal Error Rate (EER) estimated in the ROC analysis using the MODIS 500 as reference (see Sec. 8.1.2 for a more detailed definition of the EER).

While the option 'by area matching' requires only spatial consistency at level of scene, the option 'by ROC optimisation' would theoretically require spatial consistency at least better than half of the reference pixel size, i.e. 250 meters in the case of MODIS. In both cases the assumption is that the majority of the information in the scene is correctly represented by the reference layer: nevertheless, the method is robust against large deviations of the reference, empirically estimated in the order of 30% of error admitted.

Fig. 7.5 MODIS 'urban areas' used for training in the same input image of Fig. 7.1.

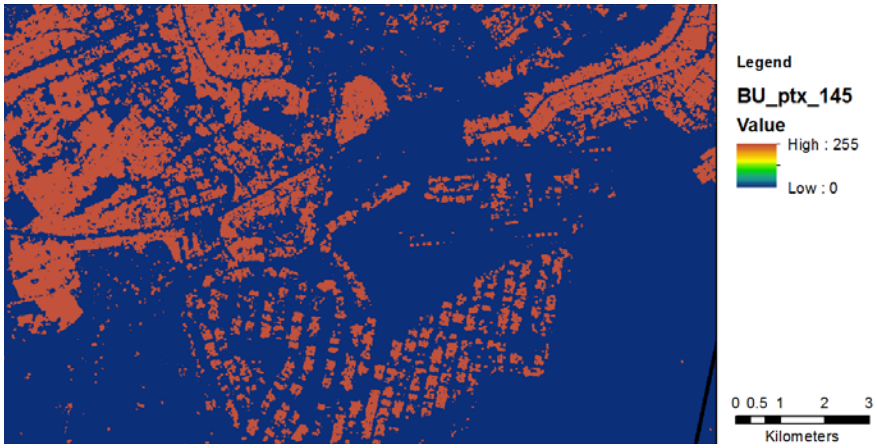


Fig. 7.6 Thresholding of the PanTex feature (Fig. 7.2) using the 'area matching' option and MODIS urban areas as reference (Fig. 7.5).

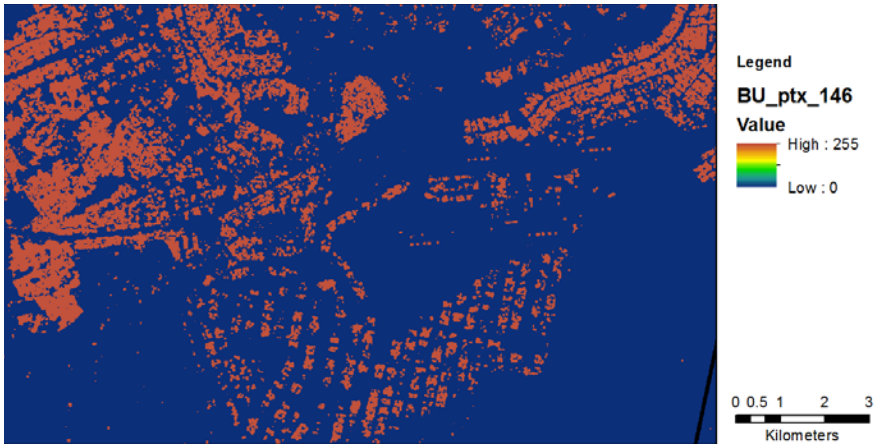


Fig. 7.7 Thresholding of the PanTex feature (Fig. 7.2) using the 'ROC optimisation' option and MODIS urban areas as reference (Fig. 7.5). This approach tends to be more conservative than the learning 'by area matching'.

7.3 Adaptive learning on morphological image features

The morphological characteristics of the image information used in the workflow are formalised through the C-Saliency-Level (CSL) model, which stores the multiscale morphological image decomposition in a compact way.

The 'C' or 'characteristic' layer of CSL includes the scale and the (opening, closing) domain of morphological response [POG12]. In other words: it contains the best estimation of the size of the image structures and their behaviour (brighter, darker) with respect to the relative background. It is a double scale-space decomposition with respect to the original image and their inverse. This layer is invariant to multiplicative and additive transformations of the input image data, and doesn't require any standardisation steps, even if used with non-calibrated, heterogeneous image data input.

The 'S' or 'saliency' layer of the CSL model instead reports about the amount of contrast explained by the specific scale transformation collected by the characteristic 'C' layer (Fig. 7.8). In other words, it is the amount of confidence —based on the available contrast between the image structure and their background— that can be given to the specific scale inferred by the CSL model for a given structure. From another point of view, this saliency can be interpreted as a measure of the fitting of the image structures with respect to a specific image information query defined by a range of scales of the target structures. In this sense, the saliency plays an important role in driving the integration between morphological descriptors and other image features for optimisation of the image information discrimination.

The L or 'levelling' layer of the CSL model was designed in order to keep the radiometric information of the image: the levelling stores the residual contrast of the image not explained by the range of scales included in the morphological decomposition. By composition of the levelling and the other two CSL layers, it is possible to reconstruct an approximation of the original input image without having access to the whole original multiscale decomposition. Being designed for exploitation of

the radiometric image descriptors, the levelling is not used in this study, where for the moment such descriptors are not exploited for discrimination of BU/NBU areas.

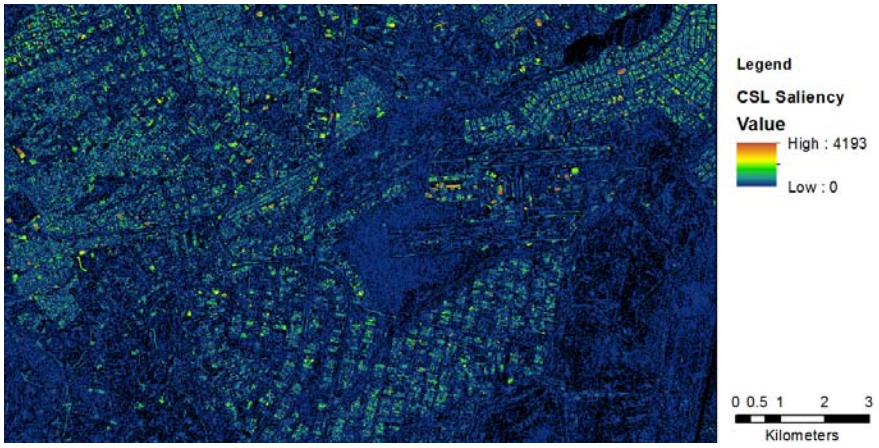


Fig. 7.8 'Saliency' feature from the CSL multi-scale morphological decomposition model derived from the input image in Fig. 7.1.

Since the saliency is linked to the contrast of the image (the levelling is not used here), it is obviously very sensitive to multiplicative transforms of the input data. Consequently, if calculated from non-calibrated heterogeneous image data input, the saliency needs to be standardised before integration with the other information layers. This standardisation is achieved through the adaptive learning described here. Several options are available in I2Q GHSL workflow parameter sets, which combine available reference layers at different scales and extrapolated typical saliency behaviours in BU and NBU reference areas. During this study those options were not benchmarked and consequently they are not discussed here. In this study, the adaptive learning of the saliency layer was performed by observing the average and standard deviation values of the saliency in the image defined compared to the BU reference areas, where available (Fig. 7.9). In the absence of reference information, which is typically the case in the 'discovery' learning phase, the BU reference area is inferred from the PanTex output that was calculated previously. In this case, an internal retroactive mechanism between textural and morphological image information descriptors was established, while in the other cases two independent, parallel learning paths were maintained.

7.4 Information fusion

In the previous sections we have described how we optimised the output from the textural and morphological analysis through the different learning approaches. The next step is the combination of both layers in order to maximise the information content and the quality. This is obtained through information fusion of the PanTex and the saliency layer of the CSL model. Only the saliency layer is used, because it can be interpreted as a confidence measure that the structures detected are

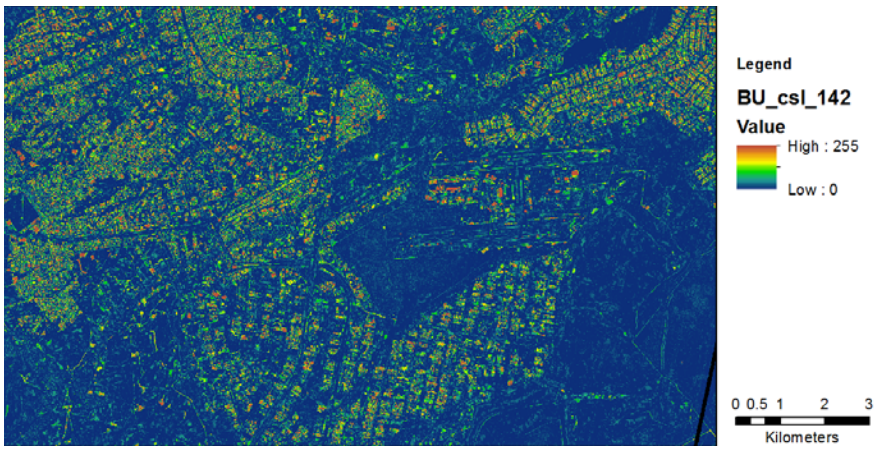


Fig. 7.9 'Saliency' feature of Fig. 7.8 rescaled after adaptive learning.

buildings. For the image fusion, three options were tested: i) intersection by closing, ii) intersection by reconstruction, and iii) simple intersection. Be the *ptx*, *sal* the PanTex and CSL saliency layers, respectively after the learning step optimising the BU/NBU recognition with a target resolution fitting the 1:10K GHSL specs, and be $SE \in [10, 50, 500]$ the structuring element with a scale corresponding to the spatial displacement admitted in the GHSL technical specs for the local, regional and global scales, respectively. The 'intersection by closing' is calculated as $y = \phi_{50}(ptx) \cap slc$, while the intersection by reconstruction is calculated as $y = \texttt{recon}(slc, \texttt{marker} = ptx)$. The 'direct intersection' option is given by pointwise function $y = ptx \cap slc$.

Figure 7.10 shows an example of input image, while Figs. 7.11 and 7.12 show the image-derived information about presence of built-up areas using, respectively, morphological and textural features after the adaptive learning phase. Figures 7.13, 7.14, and 7.15 show the output of the fusion adopting the 'direct intersection', the 'intersection by reconstruction' and the 'intersection by closing' methods. The visual comparison of the fusion results highlight the different characteristics of the approaches. The most conservative approach is the point-based direct intersection. However, this approach tends to underestimate the larger industrial buildings present in the example. This is mitigated by using the morphological operators 'by closing' and 'by reconstruction'. The intersection by reconstruction has the tendency to reconstruct also other features such as road segments.

7.5 Multi-scale generalisation protocol

The multi-scale generalisation protocol is used by the I2Q system to manage the trade-off between the precision and the computational cost of the multi-scale information representation, from local to global spatial units. While the most accurate way to summarise from local to global scales is often an aggregation based on local statistics (mean, stdv), this may have a prohibitive computational cost, if applied to global high-resolution datasets managed by standard WMS/TMS tech-

Fig. 7.10 Example of input image data (CBERS 2.5m).

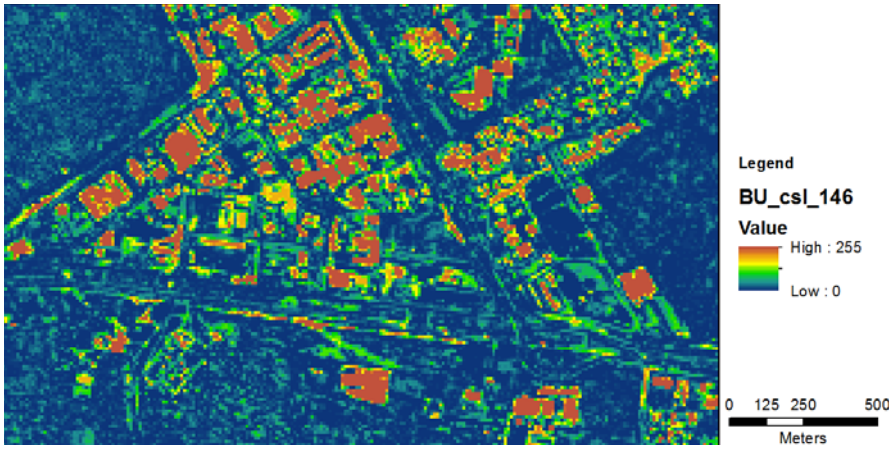


Fig. 7.11 BU areas described by morphological image information: rescaling of CSL 'saliency' image feature after the learning process. Area is identical with Fig. 7.10.

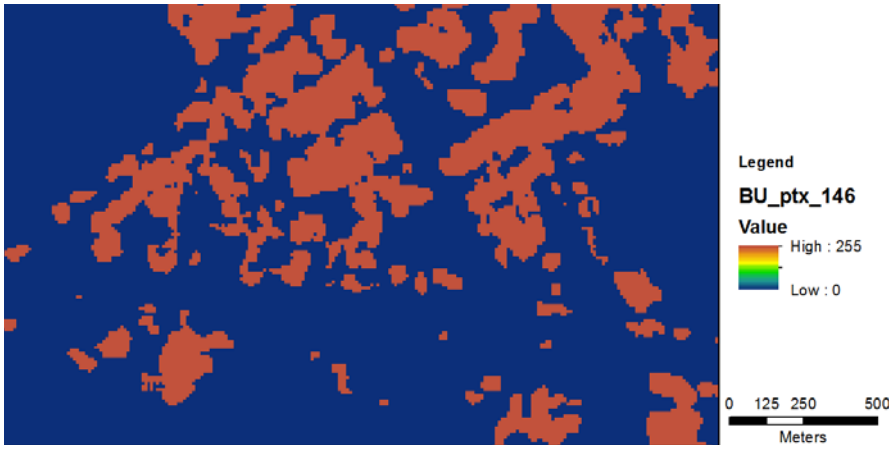


Fig. 7.12 BU areas according to textural image information: rescaling of the PanTex image feature after the learning process. Area is identical with Fig. 7.10.

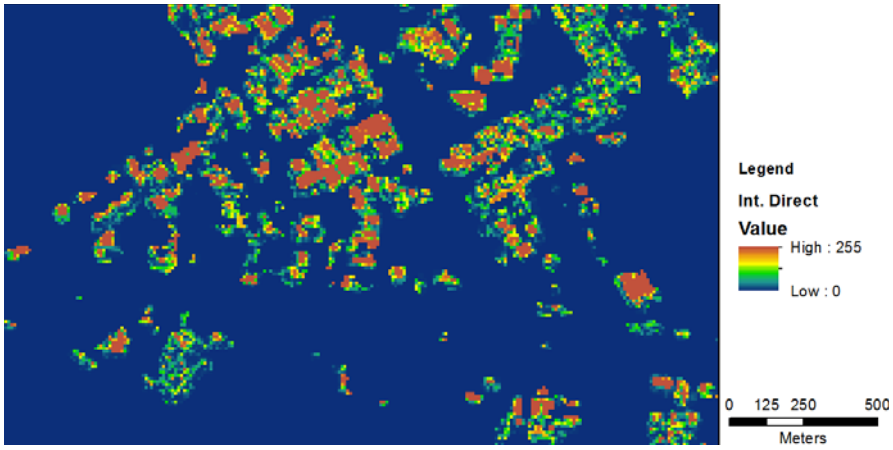


Fig. 7.13 Fusion of the textural (Fig. 7.12) and morphological (Fig. 7.11) image information using the 'direct intersection' method.

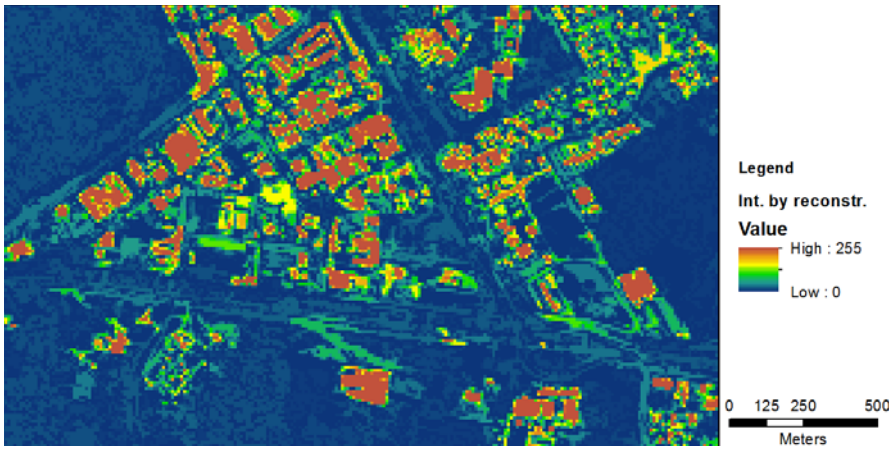


Fig. 7.14 Fusion of the textural (Fig. 7.12) and morphological (Fig. 7.11) image information using the 'intersection by reconstruction' method.

nologies and protocols. In order to maintain the interactivity of the system, the output geo-information at the local scale is prepared already at different successive generalisation scales compacted in the same layer by the means of the multi-scale generalisation protocol. In this way, the compact layer can be spatially queried at different scales, by using fast computational resampling algorithms as for example the 'nearest neighbour' algorithm. This will provide a fast approximation of computationally more expensive and more precise aggregation options. Three options are implemented in the IQ system for handling spatial generalisation i) by dilation, ii) by closing, and iii) by a hybrid approach. If IQ_{BU10K} is the result of the image information query with a target resolution fitting the 1:10K GHSL specs, and $SE \in [10, 50, 500]$ is the structuring element with a scale corresponding to the spatial displacement admitted in the GHSL technical specifications for the local, regional and global scales, the options can be formalised as follows:

- by dilation:

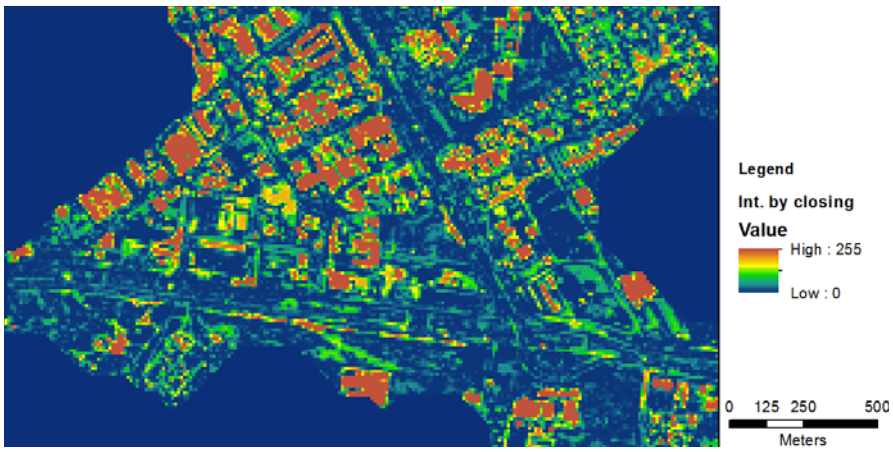


Fig. 7.15 Fusion of the textural (Fig. 7.12) and morphological (Fig. 7.11) image information using the 'intersection by closing' method.

$IQ_{BU50K} = \delta_{SE(50K)}(IQ_{BU10K})$,
$IQ_{BU500K} = \delta_{SE(500K)}(IQ_{BU10K})$;

- by closing:
  $IQ_{BU50K} = \phi_{SE(50K)}(IQ_{BU10K})$,
  $IQ_{BU500K} = \phi_{SE(500K)}(IQ_{BU10K})$;
- by hybrid approach:
  $IQ_{BU50K} = \delta_{SE(50K)}(IQ_{BU10K})$,
  $IQ_{BU500K} = \phi_{SE(500K)}(IQ_{BU50K})$.

In Fig.7.16 we show an example for the city of Sanaa, Yemen, used to test the generalisation and multi-scale composition options discussed here.

In Fig.7.17, we demonstrate the effect of the different generalisation and multi-scale composition options on the global representation. Fig.7.17.a) shows the 'best' representation of the city of Sanaa, by aggregating building footprints from local to global scale using local average operator. Fig.7.17.b) shows the global scale by using nearest neighbour resampling of the building footprints at scale 1:10K. Note how information is degraded especially in the small and scattered built-up structures and urban fringes. Fig.7.17c,d) show the same global-scale representation of Sanaa, generated by nearest neighbour resampling of the building footprints processed with the proposed multi-scale generalisation protocol. Note how these representations match better the "reference" made by local average (Fig.7.17.a), while using much faster spatial resampling techniques.

Based on the above observations the hybrid approach was used, because it is providing the best compromise for regional and global scales. Because of the needs of compression of the output mosaic and simplification of the WMS infrastructure delivering the final information during the prototype design, the three GHSL scales are fused in one only layer. The fusion is performed by the summing the three scales so that $BU_{out} = BU_{10K} + BU_{50K} + BU_{500K}$.

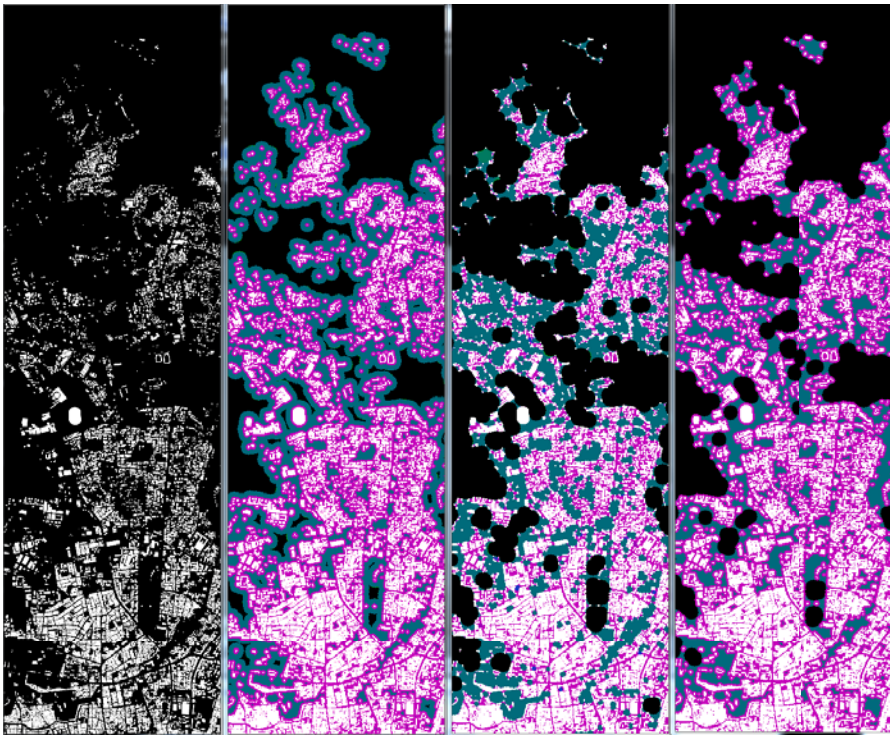


Fig. 7.16 Test of the generalisation and multi-scale composition options in the city of Sanaa, Yemen. From left to right: a) in white building footprints at 1:10K scale, generalisation by b) dilation, c) closing, and d) hybrid approach. White, pink and green, respectively, show the contribution of the 'local', 'regional', and 'global' scales to the final GHSL product.

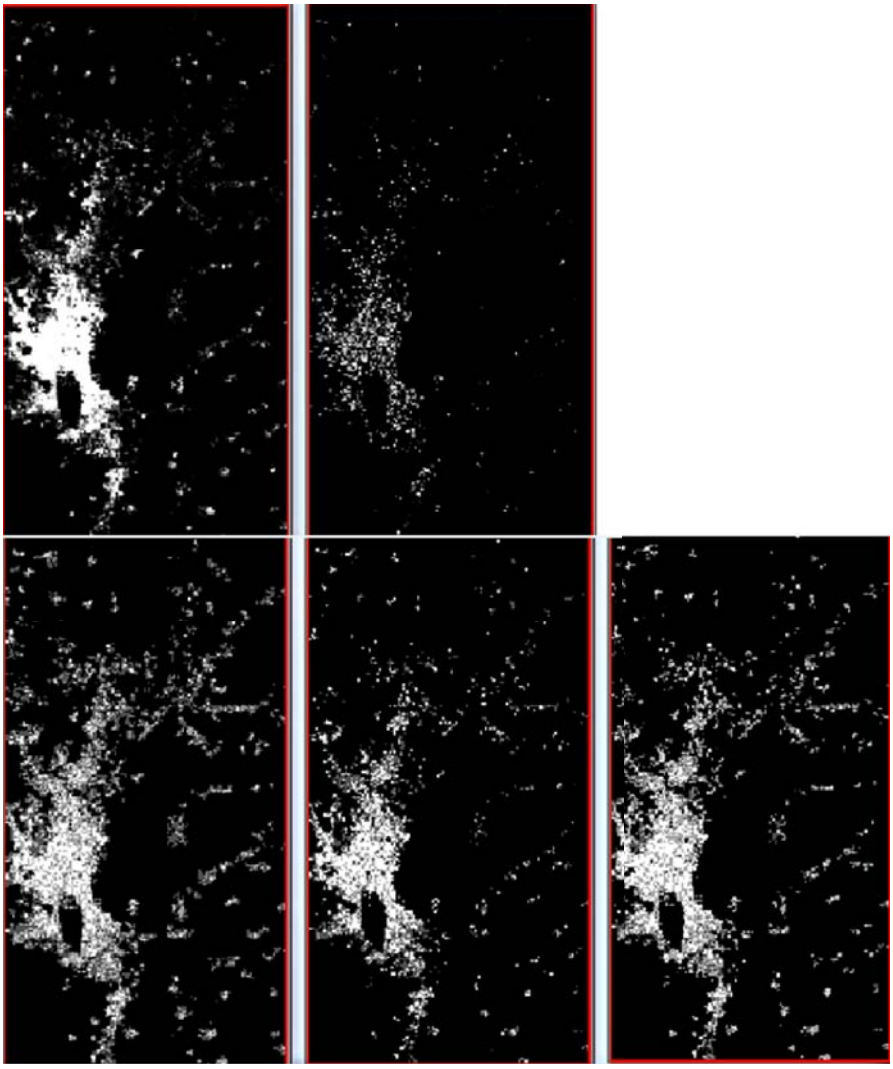


Fig. 7.17 Test of the effect in the global representation of the different generalisation and multi-scale composition options of the GHSL outputs in the city of Sanaa, Yemen. From left to right: a) building footprints from cadastral maps 1:10K aggregated to the global GHSL scale using local average; b) the same using fast "nearest neighbour" resampling algorithm, c) using "by dilation", d) "by opening" and e) "hybrid" generalisation options and also fast "nearest neighbour" resampling algorithm.

Chapter 8
Quality control and validation

This chapter describes the quality control and validation approach adopted in the context of the GHSL. The approach relies on the combination of visual collection of a limited number of reference data and globally available reference data sets. Section 8.1 describes the visual collection of reference data and the quality measures used, while section 8.2 describes the quality measurements and section 8.3 presents their assessment.

8.1 Visual validation protocol and quality measurements

The validation comprises the collection of reference data by visual interpretation and the quality assessment using well defined quality measures.

8.1.1 Visual reference data collection

In line with the focus of this study, the visual reference data collection provides dichotomic information on the presence/absence of buildings for the GHSL 50K scale. The reference data collection includes the following steps [Ehr+12]: i) collection of spatial samples by a systematic grid procedure and ii) interpretation of each sample by visual inspection of the corresponding part of the image. The collection of samples uses TMS level 17 to locate the samples. The samples are referred herein as sample blocks and outline the area on the image that will be photo-interpreted. TMS-Level 17 grid cell covers approximately 300 x 300m of the Earth's surface. Sample blocks are selected from the TMS level 17 grids using a stratified systematic sampling strategy [Gal05]. Within every sample block, 16 photo-interpretation cells are selected at TMS Level 19 (approx. 75 by 75m) for photointerpretation. Fig. 8.1 provides an example of a sampling block with the interpretation cells. Given the adopted dichotomic protocol, the interpreter is asked to check, if the specific sub-sample was intersecting a visible building in the image with four possible coded answers: `yes`, `no`, `I'm not sure`, and `no data available`. Cross-comparisons of parallel validation campaigns done by different operators on the same set of images were used in order to control the consistency and reliability of the human interpretation task.

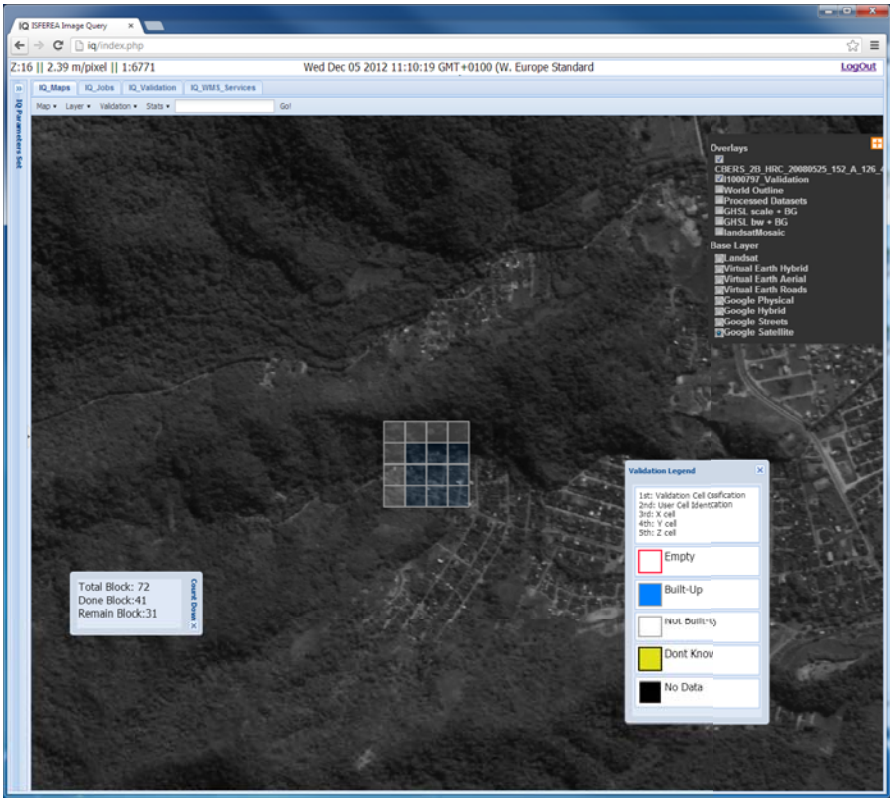


Fig. 8.1 Typical sample block with the 16 photo-interpretation cells. In this example 9 out of 16 cells have been marked as built-up. The reference data collection workflow is fully integrated in the web-based IQ system.

8.1.2 Quality Measurements

The quality measurements are used to provide standardised, automated descriptors of the quality of the processing output compared to a reference. Given the continuous measurements $\{x_i\}$ and the classes $\}, y_i \in \{NBU, BU\}$ that the measurements should predict, their classification power is best captured by the Receiver Operational Characteristics (ROC) [Kul68]. The ROC analysis was developed in signal detection theory. It is a graphical plot which illustrates the performance of a binary classifier system. It is created by plotting the fraction of true positives out of the positives against the fraction of false positives out of the negatives. In other words the ROC curve represents the probabilities of missed detection and false alarm for various classification thresholds:

$$p_{md}(T) = \frac{\sum_i 1(x_i < T)1(y_i = BU)}{\sum_i 1(y_i = BU)} \tag{8.1}$$

$$p_{fa}(T) = \frac{\sum_i 1(x_i \geq T)1(y_i = NBU)}{\sum_i 1(y_i = NBU)} \tag{8.2}$$

The ROC curve is the convex parametric function $(p_{md}(T), p_{fa}(T)), \forall T$. In order to compare two types of measurement producing two ROC curves, the metrics of the Area Under the Curve (AUC), the Equal Error Rate (EER), and the Minimal Error Rate (MER) are proposed and can be interpreted as quality statistics:

$$AUC = \int_0^1 p_{fa} dp_{md} \tag{8.3}$$

$$EER = p_{md}(T_e) = p_{fa}(T_e) \exists T_e \tag{8.4}$$

$$MER = min\left[p_{md}(T_e) + p_{fa}(T_e)\right], \exists T_e \tag{8.5}$$

The smaller these quality metrics are, the better is the classification or discrimination performance of the type of measurement.

8.2 Global quality measurements

While a visual validation protocol is highly accurate, it lacks completeness. With the adopted protocol, the complete collection of for each satellite scene used in the experiment would require several years of visual reference data collection for an image interpreter, which is totally unrealistic with respect to the available allocated resources. Despite this fact, a systematic evaluation of the results of the automatic image information retrieval is needed in order to control the consistency of the output having multiple heterogeneous input scenes. This is crucial in particular for the generation of the final mosaic, for which partially overlapping scenes with possibly different quality outputs have to be managed in order to maximise the quality the final mosaic.

To alleviate this issue, we propose the consideration of a global/complete layer representing the presence of 'built-up' at 1 km resolution (BUref). This layer is derived by the combination of available global information layers reporting about the presence of buildings. In the specific case we adopted a combination of MODIS and LandScan. This combination is obtained by reclassification of LandScan using MODIS as training set, with a spatial learning kernel of 500 kilometres. As this reference is coarser than the visual reference derived in the GHSL 50K validation protocol, the errors are two-fold: they are either true or they are due to the mismatch of spatial resolution. Because of these effects the average accuracy computed with respect to this reference is less reliable. Nevertheless, we use this quality metric because it can be computed globally. Various communities faced the problem of validating automatically any result of a processing flow. As an example, the JPEG/JPEG2000 community introduced objective quality metrics of compressed images, which are the best correlated to subjective quality metrics obtained from feedback from a pool of persons [Wan+04]. Similarly, we propose to link the global quality assessments with the with the visual ones. In the next section we analyse the relationship between the visual reference data and the global reference data sets, which will provide an insight in the linkages of the different data sets, which is important to evaluate properly the results of the cross-comparison between GHSL output and the global data sets.

8.3 Assessment of quality measures

In this section, we investigate two issues: 1) what is the appropriate quality measure for evaluating the accuracy of the GHSL output against available reference data?, and 2) admitting that human-based visual interpretation is an expensive process, could a performance assessment that is based on reference data provided by MODIS 500-m and LandScan exhibit high confidence and reliability?

For the needs of this assessment a total of approximately 95,000 BU and 700,000 BNU samples were visually collected as described above. The samples were collected mainly from 2.5m resolution, pansharpened SPOT-5 and 0.5m resolution aerial RGB imagery. The total extent covered by this visual interpretation protocol was over 700,000km^2.

To provide an answer for the first question, we deploy a ranking schema based on regression analysis as explained below: First, we calculate the ROC-derived quality measures of equations (8.3), (8.4) and (8.5) for each of the following pairs: GHSL outcome against BUref, GHSL outcome against MODIS, PanTex against BUref, PanTex against MODIS. This results in $3 \times 4 = 12$ quality measures. We include also PanTex in this assessment in order to check if the, theoretically information enriched, GHSL outcome remains in practice more informative than PanTex.

Next, by seeking correlations among the 12 quality measures and the three quality measures of equations (8.3), (8.4) and (8.5) calculated on the pair GHSL outcome against results by visual interpretation, we apply polynomial curve fitting of various degrees in a least squares sense. As goodness-of-fit, for each curve we compute (at 95% confidence interval):

(a) the correlation coefficient R,
(b) the coefficient of determination: $R^2 = 1 - \frac{RSS}{TSS}$,
(c) the adjusted $R^2_{adj} = 1 - \frac{(n-1)RSS}{(n-k-1)TSS}$,
(d) the Mean Squared Error: $MSE = \frac{1}{n}RSS$, and
(e) the Bayesian Information Criterion: $BIC = n \ln\left(\frac{RSS}{n-k-1}\right) + k \ln(n)$

where $RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$, $TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$, n is the total number of observations, k is the number of model parameters, y are the values of the 12 quality measures described in the previous step, $\bar{y}$ is the average value of ys for each of the quality measures, and $\hat{y}$ denotes the fitting values (the response of the regression models). The regressors are the three quality measures (8.3), (8.4) and (8.5) calculated on the pair GHSL outcome against results by visual interpretation. When comparing multiple models, big values of the criteria (a), (b) and (c) are better, while the opposite holds for the criteria (d) and (e), i.e. the smaller the better.

In the general case, it is possible to experiment with p_1 numbers of parametrisation x p_2 quality measures x p_3 pairs of system outputs against reference data x p_2 quality measures x p_4 pairs of system output against visual interpretation x p_5 number of regression models.

In our study, we experiment with polynomials of the form $y(x) = c_0 x^{k-1} + c_1 x^{k-2} + \ldots + c_{k-1}x + c_k$, where the integer k takes values in the range $[2, 11]$. In addition to keeping $k = 2$ we made tests with generalised non-linear models like exponential, logarithmic and power law models. This results in 13 models x 3^2 quality measures x 4 pairs of variables $= 468$ figures. Due to the fact that we tested

and applied three different parameter sets, our regression analysis was based in total on $1,404$ figures.

At the final stage, we decided to rank the regression models in ascending order of both MSE and BIC. The MSE criterion describes for the accuracy between the expected and the true value by quantifying their difference; the BIC criterion is used to prevent that the model is overfitting by penalising the increase on the number of parameters in the model. According to these criteria, the best ranking is the polynomial model of 2^{nd} order that fits the MER of GHSL outcome against visual interpretation and the MER of GHSL outcome against BUref. It results in the $MSE = 0.3892$ and the $BIC = -170.6723$.

In Fig. 8.2 we display the curves fitting with respect to the three quality measures AUC, EER and MER that achieve the best adaptation according to the defined goodness-of-fit criteria. The first one plots the AUC between the outcome of GHSL and visual reference against the EER between the outcome of GHSL and BUref
(BUfinal.visual.AreaUnderCurve against BUfinal.ROC|BUref.equalErrorRate)
giving $R = 0.460$, $R^2 = 0.212$, $R^2_{adj} = 0.208$, $MSE = 0.784$ and $BIC = -35.500$. The middle one plots the EER between the outcome of GHSL and visual reference against the EER between the outcome of GHSL and BUref
(BUfinal.visual.equalErrorRate against BUfinal.ROC|BUref.equalErrorRate)
giving $R = 0.376$, $R^2 = 0.245$, $R^2_{adj} = 0.233$, $MSE = 0.750$ and $BIC = -31.562$. The third one plots the MER between the outcome of GHSL and visual reference against the MER between the outcome of GHSL and BUref
(BUfinal.visual.minimalErrorRate against BUfinal.ROC|BUref.minimalErrorRate)
giving $R = 0.745$, $R^2 = 0.608$, $R^2_{adj} = 0.604$, $MSE = 0.389$ and $BIC = -170.672$. The code word "BUfinal" employed in the names of the previous variables denotes the GHSL outcome.

Figure 8.3 displays instances of the layers under discussion (from left to right and from first row to the second one): visual interpretation (sparse sampling via GHSL 50K protocol), MODIS 500-m, BUref and GHSL outcome.

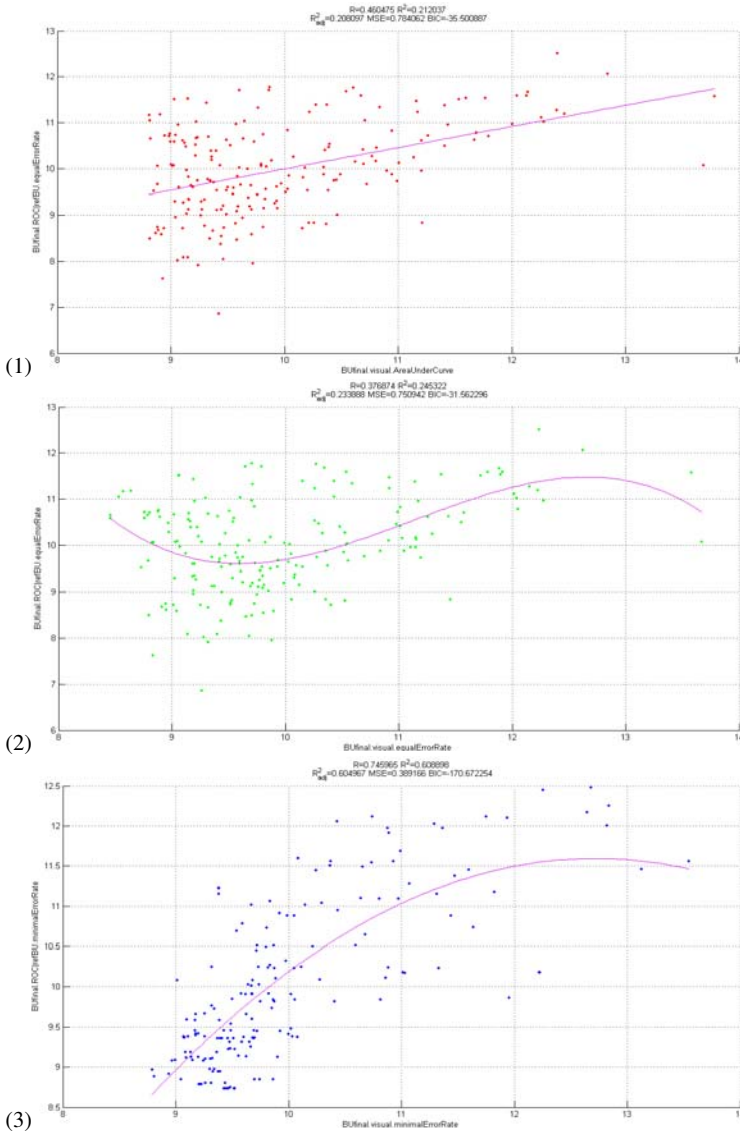


Fig. 8.2 Representative plots created through regression analysis. From the first to the third, they display the best result according to the chosen fitting criteria for the three quality measures AUC, EER and MER respectively. In all three cases, the better fitting was observed in the couple GHSL outcome against visual interpretation combined with GHSL outcome against BUref.

The described ranking schema leads us to observe that there is a high positive correlation (almost linear) between the *Minimal Error Rate* calculated over the pair GHSL outcome with visual interpretation, and the one computed over GHSL outcome and BUref layer. This provides a significant evidence that this particular quality measure is consistent in relation to reference data derived either by human visual inspection, or by layers of type MODIS 500-m or BUref. Following ROC curve analysis to assess the classification performance, MER provides an informa-

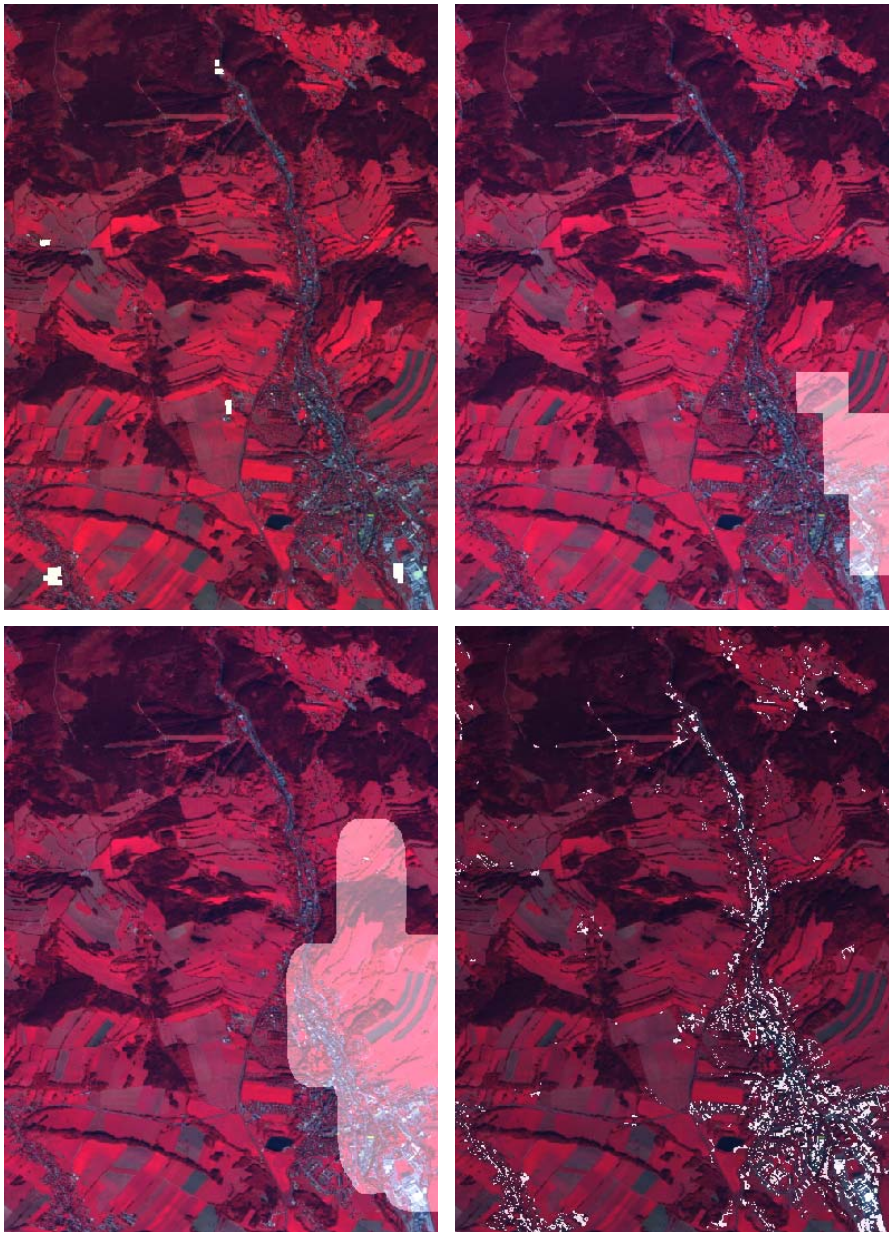


Fig. 8.3 The background image is a pansharpened SPOT-5 image with $2.5 \times 2.5m$ pixel size (Jelenia Gora, Poland). From left to right and from first row to the second one, the white overlays derive from: visual interpretation (sparse sampling via GHSL 50K protocol), MODIS 500-m, BUref and GHSL outcome. All images are resized in $10m$ resolution.

tive means to compare a BU/NBU layer originated by HR/VHR imagery with low-resolution global reference data.

Concerning the second issue posed at the beginning of this section, we test the following hypothesis: by inverting the relationship between response and regressors in the dominant polynomial model and inserting as input the MER calculated over

GHSL outcome and BUref layer (for a total surface of extent more than 24 millions km^2), could the extrapolated values of MER over GHSL outcome and visual interpretation layer be within acceptable confidence intervals?

The execution of a two-sample Kolmogorov-Smirnov test showed that the distribution of extrapolated values and the sampling distribution of MER over GHSL outcome and visual interpretation are not the same continuous distribution. This seems to be logical since the size of the sampling distribution and the size of the extrapolation is very asymmetric in favour of the latter. Nevertheless, both distributions can be approximated well by the exponential distribution giving means and standard deviation 1.5 ± 0.21 (sampling) and 1.4774 ± 0.30 (extrapolated). The sampling distribution encompasses greater variance, probably due to its relatively limited size; until we have expanded more the validated cases provided by human inspection, it will remain questionable whether a quality measure such as MER that is based on low-resolution global reference data could potentially replace quality measures calculated over the results of visual interpretation.

Chapter 9
Global results

This chapter describes the results of the processing starting with the qualitative assessment in the next section. It is followed by a quantitative analysis, first against the visual reference data and, second, against global data sets addressing different aspects such as quality by sensor, spectral bands, spatial resolution and by ecoregions of the world.

9.1 Qualitative analysis

Figure 9.1 illustrates a typical example of the GHSL output extracted from the remote-sensing images . For the same area also the MODIS500 and the LandScan population is illustrated that are used as a reference for visual comparison. The top-left image shows the 'presence of buildings' layer of the GHSL over the city of Brasilia. It is produced from image data of the CBERS-2B sensor. The output image is shown at a 1:50K 'regional' scale. The pixel brightness is proportional to the percentage of built-up presence in the specific spatial units. The dark-green polygonal contours correspond to the footprints of satellite scenes that were available as input for the final mosaic. Only the best information extracted from all the available processed scenes is used for each spatial unit. The top-right image shows a zoom into the central region of the city. The image shows the 'average building size' GHSL information layer at a 1:10K 'local' scale . The colour-coding follows the blue-green-yellow-red order on increasing size of the built-up structures.

Comparing the output of the GHSL produced from HR image data against the other two low resolution layers, a noticeable gain is observed with respect to the detail of the extracted information. Moreover, the quality of the GHSL scenes offers the option of addressing the internal characterisation of built-up areas by automatic analysis of the morphological characteristics of the built-up structures. Recognition and characterisation of single built-up structures is evidently not possible using low-resolution input images.

Figures 9.2, 9.3, and 9.4 show image data, GHSL output and MODIS output, respectively, in the region around the town of Tongeren, Belgium, showing scattered medium-size settlement patches. The strong underestimation of these kind of patterns in the low-resolution MODIS urban layer is evident. In this area, the GHSL information was produced using 2.5-m resolution pan-sharpened SPOT image data.

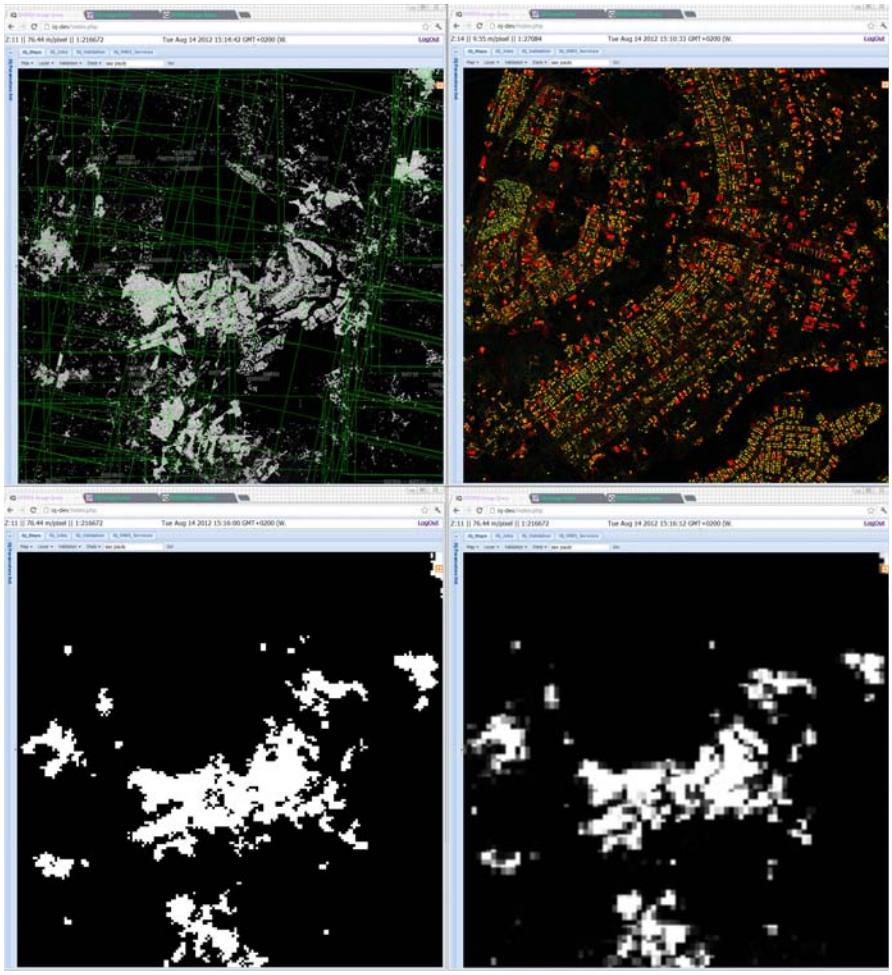


Fig. 9.1 City of Brasilia. Top-left: the 'presence of buildings' GHSL layer represented at 1:50K scale with the footprints of input scenes (CBERS-2B) in dark-green; top-right: a zoom into the city centre. The image shows the 'average building size' GHSL layer at 1:10K scale. Increasing built-up size is mapped on the blue-green-yellow-red colour map; bottom-left and right: the same city represented by the MODIS500 urban layer and BUref respectively.

Figures 9.5, 9.6, and 9.7 show image data, GHSL output and MODIS output, respectively, in the region around the Gaza strip that is one of the highest density population areas of the world. The input images used for the GHSL production in this areas are produced by a mixture of different sensors ranging from 0.5 to 2.5 meter of spatial resolution. In Fig. 9.6 the colour encodes the estimated size (surface) of built-up structures in the range of 10–15,000 square meters from blue (small) to red (large). Also in this case, the strong bias introduced by the low-resolution MODIS sensor in detecting medium and small size settlement patches is noticeable.

Figures 9.8 , 9.9, and 9.10 show the city of London (UK) at three different scales, namely 1:500K, 1:200K and 1:100K. They represent the GHSL output reporting the estimated size (surface) of built-up structures. The colour encodes the size information in the range of 10–15,000 square meters from blue (small) to red (large). It

Fig. 9.2 Image data (mosaic of Landsat ETM) collected over the region around the city of Tongeren (Belgium).

is interesting to explore the relations between different colour patches in this cartographic representation and in the geographical domain. Visual checks reveal that there is an evident spatial correlation between the presence of 'red' patches and the presence of industrial/commercial buildings as opposed to residential areas appearing mostly as bluish patterns. Yellow-orange patterns are instead correlated to central areas. These observations remain only at the qualitative level and indicate the potential use of GHSL image-derived measurements for internal settlement pattern characterisation. More systematic assessment would need the collection of suitable reference data and the definition of a formalised validation protocol that is not included in the current report.

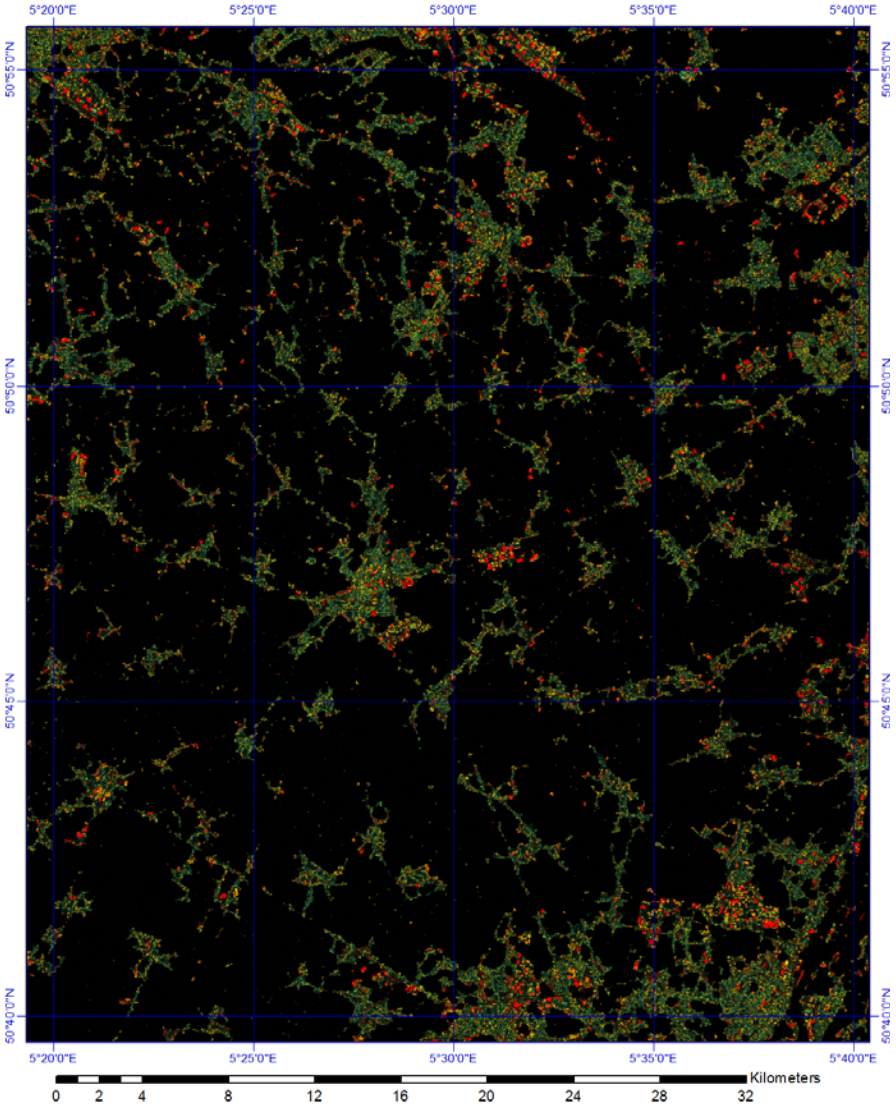


Fig. 9.3 GHSL output over the same area of Fig. 9.2. In the image the colour encodes the estimated size (surface) of built-up structures in the range of 10–15,000 square meters from blue (small) to red (large).

9.2 Quantitative analysis with visual reference data

The quantitative assessment of the quality of the GHSL output against visual reference data relies on a total of approximately 95,000 and 700,000 samples of BU and NBU classes respectively, which were collected using the GHSL reference data collection protocol (see Sec. 8.1.1. The samples were collected mainly from 2.5m resolution, pansharpened SPOT-5, CBERS-2B, and 0.5m resolution aerial RGB imagery. The total ground surface processed employing this visual interpretation protocol was over 700,000 km^2. The results of the quantitative analysis over a subset of image data, for which visual reference information was available, are presented

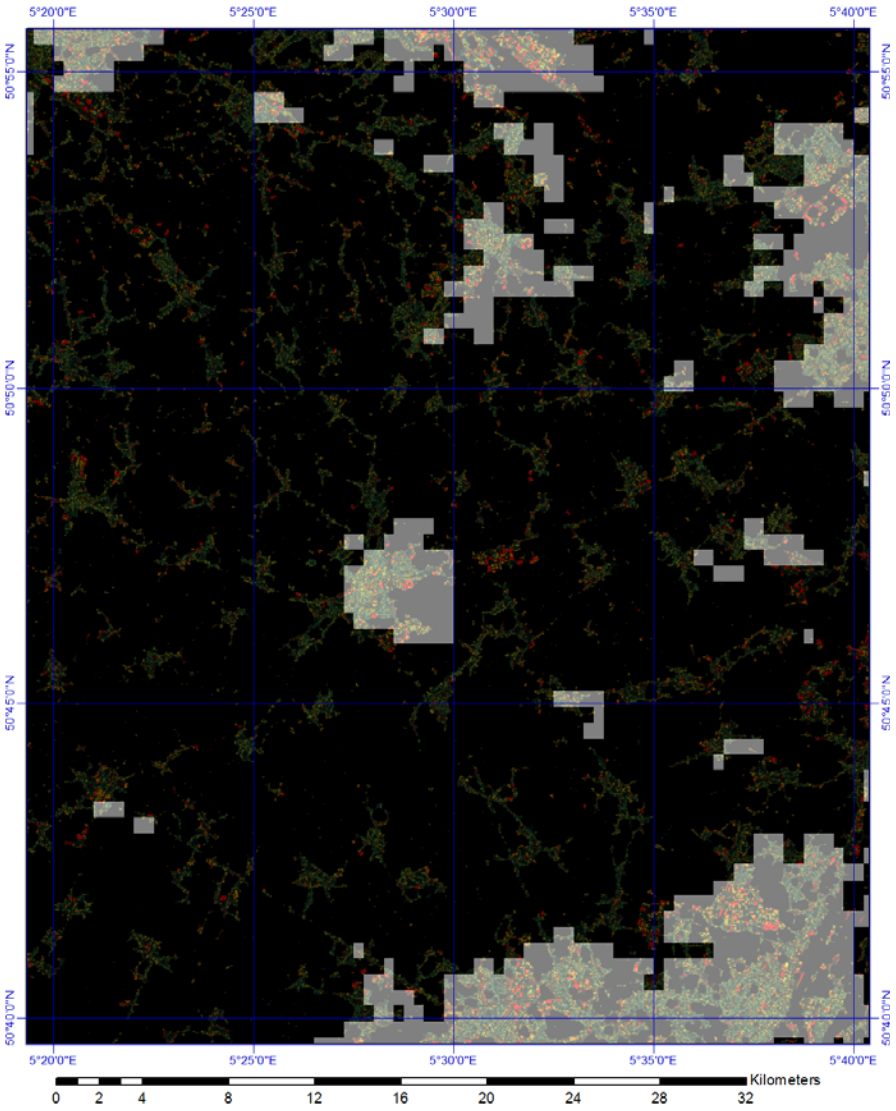


Fig. 9.4 Comparison of the GHSL and of the MODIS500 outputs in the same area of Fig. 9.2. In white the information coming from the MODIS urban areas.

in Table 9.1. The error rates are listed with respect to three different parameter sets benchmarked during this study: namely 142, 145, and 146. Each relates to a set of options activated in the learning and classification steps, and is briefly introduced in the following.

- Learning in option 142 employs the expectation-maximisation strategy using the LandScan 1km-resolution population layer as reference;
- Learning in options 145 and 146 both uses the MODIS 500m-resolution urban areas as reference following two different strategies: cumulative histogram matching (145) and minimisation of the EER (146).

Fig. 9.5 Image data representing the region around Gaza strip (mosaic of Landsat ETM)

The adopted visual validation protocol shows an estimated EER and MER of the final GHSL built-up layer of approximately 17% and 9% respectively. This corresponds to a total accuracy of more than 90%, i.e. $(1 - MER)$. All three learning-and-classification options evaluated using the proposed protocol appear to be consistent and have only minor differences in performance. Option 142 ranks best with respect to the EER metric (17.26%), and option 145 performs best with respect to the MER metric (8.57%). Option 146 has the lowest rank in both metrics.

In order to allow an assessment of the feasibility to use the global reference layers for the quality assessment, the two quality metrics were also computed for the MODIS 500m-resolution and the BUref data sets taking into account only the areas, where also visual reference information is available. for the same satellite input imagery and using the same visual validation protocol. The results are listed in Table 9.2.

Despite minor differences that are being investigated further, two main observations are made: i) the quality metrics computed using the visual protocol and using

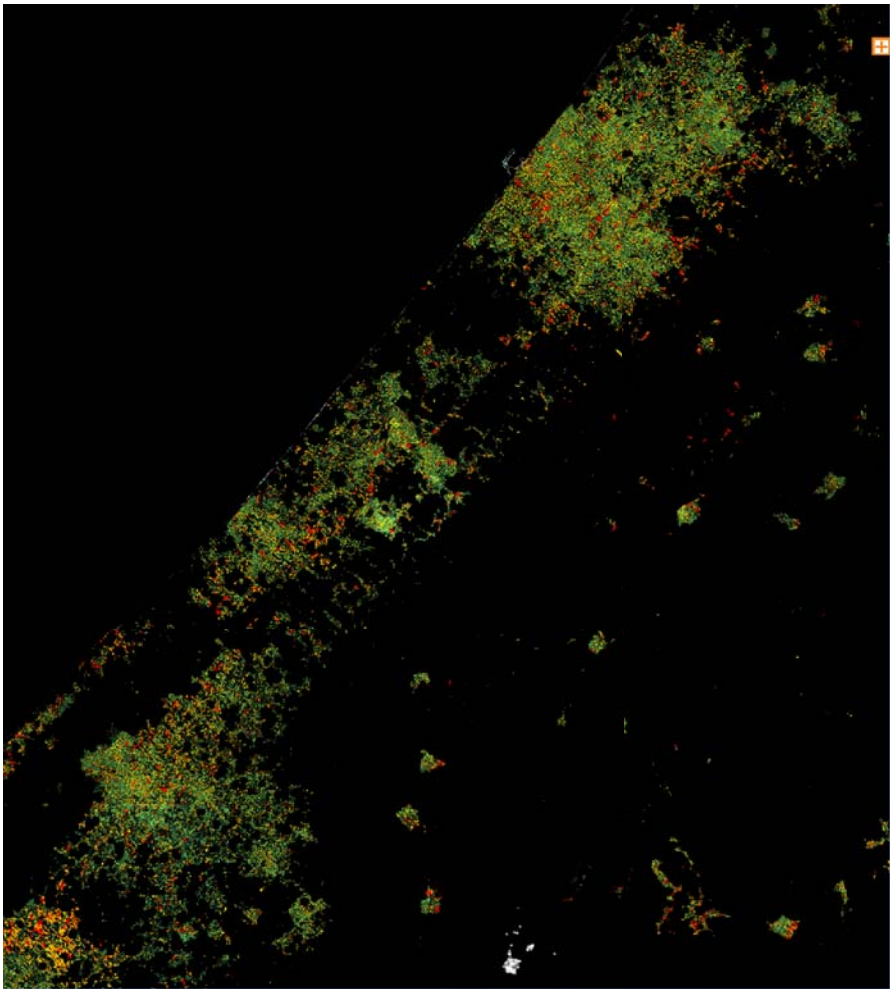


Fig. 9.6 GHSL mosaic output over the same area of Fig. 9.5 using a composition of input images ranging from 0.5 to 2.5 meters of spatial resolution. In the image the colour encodes the estimated size (surface) of built-up structures in the range of 10–15,000 square meters from blue (small) to red (large).

.

the low-resolution reference layers appear to be substantially consistent and ii) there is an almost systematic overestimation of the average error in both error metrics using the low-resolution references, if compared against the corresponding metrics using the visual references, which we assume to be more reliable than the low-resolution reference. The first observation supports the use of global, low-resolution reference layers for an automatic evaluation and ranking of the automatic image information extraction output. This is backed by obtaining almost the same ranking, if benchmarking the three learning-and-classification options on low-resolution reference layers, or if using the much more expensive visually-collected reference data.

The second observation refers to a general issue when using low-resolution reference layers for evaluating HR and VHR classification outputs. There is an error

Fig. 9.7 Comparison of the GHSL and of the MODIS500 outputs in the same area of Fig. 9.5. In white the information coming from the MODIS urban areas.

due to a mismatch of the data sets that can be due to: i) true errors; e.g. BU/NBU wrongly falling in the complementary class and correctly recognised in the low-resolution reference, and ii) false errors; e.g. BU/NBU correctly recognised at the HR, VHR output, but not found in the low-resolution reference. This second type of mismatch can be deemed 'added value' of the HR, VHR automatic recognition procedure with respect to the available low-resolution information layers.

These observations lead to two different findings: i) a positive one that stems from the fact that we can expect a substantial underestimation of the final accuracy of the GHSL layers using the low-resolution reference, meaning that the whole GHSL quality assessment can be considered as globally very conservative, and ii) a negative one that can be linked to the risk of discarding GHSL outputs showing low quality metrics, but in fact providing a dominant 'added value' with respect to the low-resolution references. In the 'discovery' phase of the learning and classification processing chain we try to mitigate this risk in the cases, where low-resolution refer-

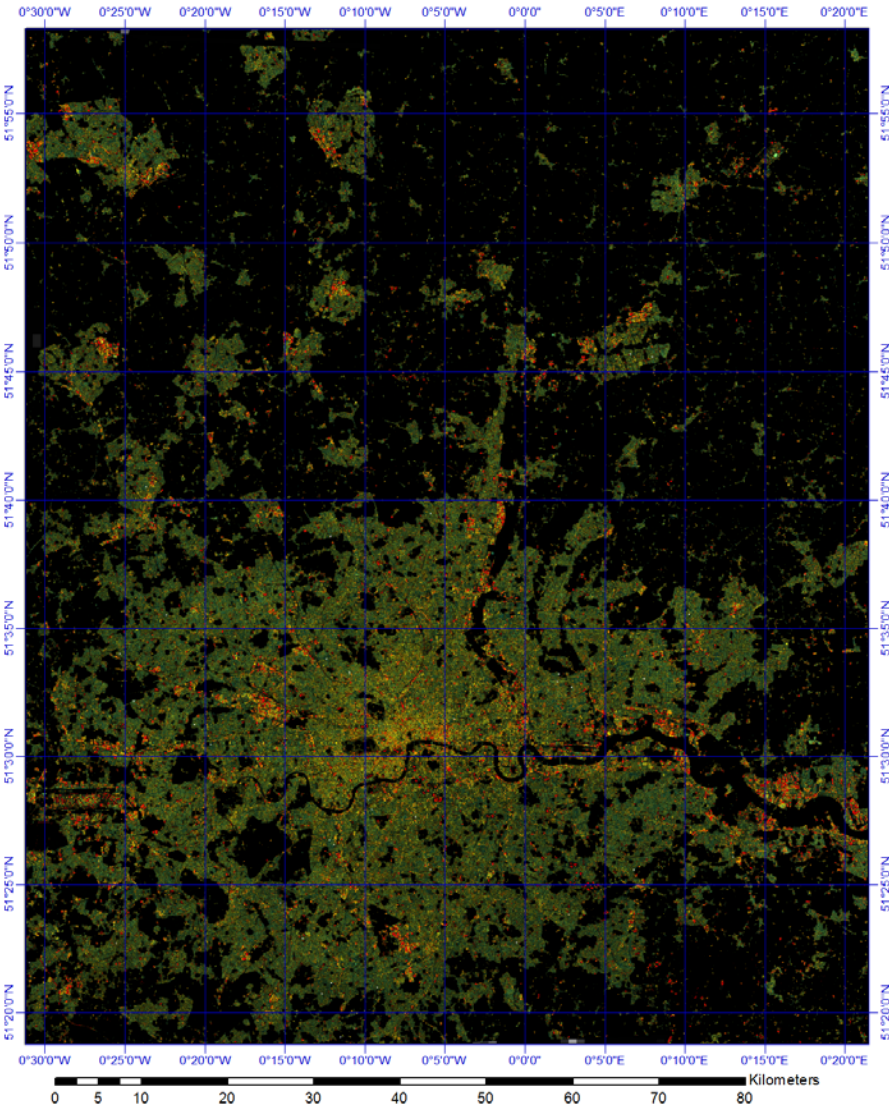


Fig. 9.8 GHSL output in the region around London city (UK) represented at the nominal scale of 1:500K. In the image the colour encodes the estimated size (surface) of built-up structures in the range of 10–15,000 square meters from blue (small) to red (large).

ences show clearly unreliable behaviour; typically in cases of low-density scattered settlements, which cannot be detected with low resolution sensors such as MODIS. The problem remains open and needs to be addressed in the next release of the learning-and-classification workflow.

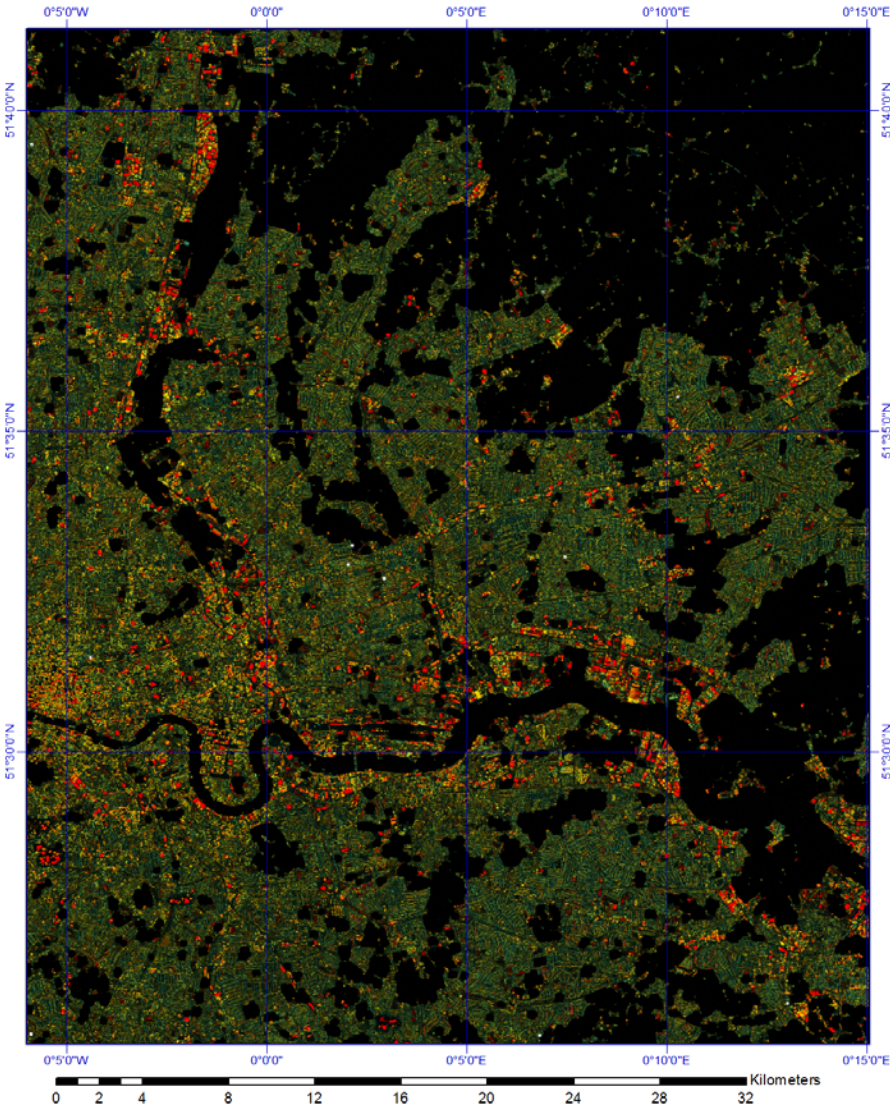


Fig. 9.9 GHSL output in the London city (UK) represented at the nominal scale of 1:200K. In the image the colour encodes the estimated size (surface) of built-up structures in the range of 10–15,000 square meters from blue (small) to red (large).

9.3 Quantitative analysis with global data sets

This section provides the estimated average quality of the whole GHSL output using the MER measure and the 1km-resolution BUref layer as reference. This is the quality metric that showed the best matching with the visual validation protocol figures. The following sections describe the analysis of the results ordered by the type of sensor, by band, and by sensor resolution, respectively. The quality metric reported in these tables is defined as the inverse of the error metric: $Q(x) = 1 - MER_{BUref}(x)$. In all three tables the results are listed by decreasing values of the $Q(x)$ metric; best ranks appear at the top of each list. Both average

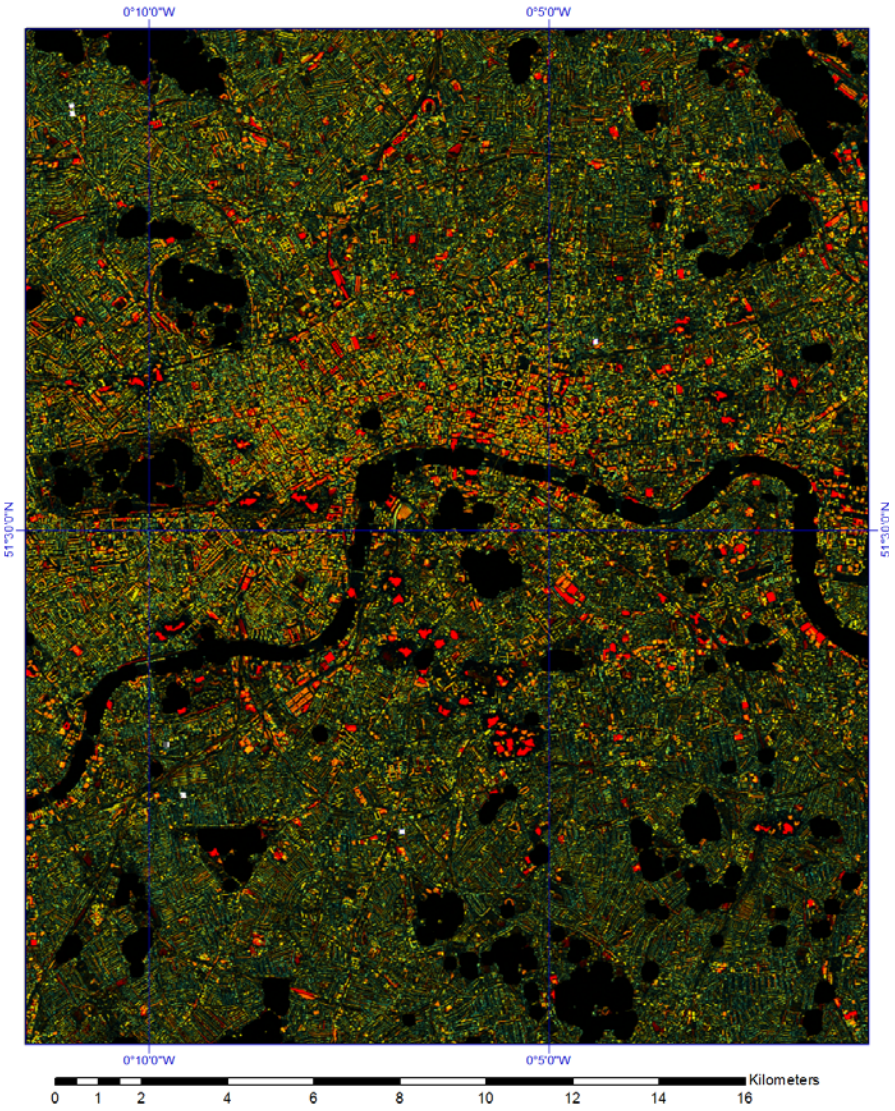


Fig. 9.10 Zoom of the GHSL output in the London city (UK) represented at the nominal scale of 1:100K. In the image the colour encodes the estimated size (surface) of built-up structures in the range of 10–15,000 square meters from blue (small) to red (large).

and standard deviation (StdDev) of the $Q(x)$ metric are shown. All values of the $Q(x)$ metric presented in these tables represent the average computed between all the available adaptive learning-classification options.

9.3.1 Quality by input sensor

The analysis of the results by sensor (Table 9.3) shows a cluster of sensors providing good BUref fitting performances with $Q(x) \geq 90\%$. These include RapidEye

Table 9.1 Error rates (EER and MER) for the GHSL output (BUfinal) and the visual reference data for each parameter set. The first rows describe the surface and the number of available reference data sets. Due to some errors in the processing of the parameter sets, there is a difference on the number of samples per set.

	Parameter_script_ID			
Data	142	145	146	Grand Total
Sum of input surface (km^2)	733,560	727,348	731,181	2,192,089
Sum of BUfinal (visual) number of positive samples	96,064	95,515	95,758	287,337
Sum of BUfinal (visual) number of negative samples	701,030	688,692	693,165	2,082,887
Average of BUfinal (visual) EER	17.26%	17.42%	18.43%	17.70%
Average of BUfinal (visual) MER	9.15%	8.57%	9.74%	9.15%

Table 9.2 Error rates (EER and MER) for the GHSL output (BUfinal) and the global reference data for each parameter set.

	Parameter_script_ID			
Data	142	145	146	Grand Total
Average of BUfinal ROC_{refBU} EER	29.84%	36.57%	37.06%	34.48%
Average of BUfinal ROC_{MODIS} EER	28.64%	30.08%	31.32%	30.01%
Average of PanTex ROC_{refBU} EER	34.23%	38.91%	38.92%	37.34%
Average of PanTex ROC_{MODIS} EER	33.04%	33.03%	33.03%	33.03%
Average of BUfinal ROC_{refBU} MER	7.93%	13.37%	13.61%	11.63%
Average of BUfinal ROC_{MODIS} MER	8.94%	9.02%	9.19%	9.05%
Average of PanTex ROC_{refBU} MER	8.17%	13.78%	13.72%	11.88%
Average of PanTex ROC_{MODIS} MER	9.26%	9.24%	9.24%	9.25%

1 and 5, aerial, CBERS-2B and SPOT-5. A medium-performance cluster of sensors with $Q(x)$ between 85% and 90% includes the SPOT-2, IKONOS-2, and RapidEye-2 sensors. A lower performing cluster of sensors, i.e. with $Q(x)$ between 80% and 85%, includes the WorldView 2 and 1, QuickBird 2, and GeoEye 1 sensors. Table 9.3 provides an interesting observation. The cluster of best-performing sensors, if VHR aerial image data is excluded, includes traditional HR platforms such as CBERS and SPOT and even the 5m resolution RapidEye sensors, while the VHR platforms, such as WorldView, QuickBird and GeoEye, rank lower. But the resolu-

tion cannot explain sufficiently the low ranking of the new VHR satellite platforms, because the aerial imagery, which was processed at 0.5m resolution, ranks second best. Evidently the proposed workflow for automatic image information retrieval is influenced by other data characteristics that remain to be further investigated for improving the overall performance. It is also important to remember that there is a strong emphasis on the HR sensors, both in terms of amount of data sets and in terms of visual reference data collection. This may well influence the quality assessments also in the lowing sections.

Table 9.3 Average quality by sensor.

	Values	
	Average of	StdDev of
Sensor	$Q(x)$	$Q(x)$
RapidEye-4	97.94%	2.27%
Aerial	95.66%	3.69%
CBERS-2B	94.37%	7.59%
SPOT-5	93.26%	7.33%
RapidEye-1	90.26%	10.10%
SPOT-2	88.21%	3.10%
IKONOS-2	86.63%	12.67%
RapidEye-2	86.38%	10.36%
WorldView-2	84.37%	10.52%
QuickBird-2	83.58%	13.04%
GeoEye-1	80.98%	11.58%
WorldView-1	80.03%	14.28%
Grand Total	91.50%	10.08%

9.3.2 Quality by input bands

Anacin the results with respect to the estimated output $Q(x)$ in relation to the different spectral bands used as input (Table 9.4), three main clusters with distinct behaviour are observed: i) a top ranking cluster with $Q(x) \geq 90\%$ including the PAN and GREEN, RED and NIR bands in pan-sharpening imaging mode (PSH); ii) a medium ranking cluster with $Q(x)$ ranging from 85% to 90% that includes all listed bands in multispectral imaging mode (MUL), and iii) and a low-ranking cluster with $Q(x)$ ranging from 75% to 80% that includes VHR bands in PSH mode. It is interesting to note that except for the BLUE band all other bands in the lowest-ranking cluster are the new bands of the WorldView-2 satellite. This behaviour of the sensor needs to be analysed in more detail. Possible explanations could be linked to the fact that there only a limited number of 20 scenes was processed that might be influenced by other characteristics. Another source of error might be linked to the pan-sharpening of these new bands. Some algorithms are not providing good results for this new band constellation.

Table 9.4 Average quality by imaging type and band.

Image type and band	Values Average of $Q(x)$	StdDev of $Q(x)$
PSH GREEN	93.56%	7.53%
PSH RED	93.54%	7.56%
PSH NIR	93.53%	7.78%
PAN PAN	93.30%	8.76%
MUL SWIR	91.33%	8.57%
MUL REDEDGE	90.18%	10.25%
MUL GREEN	86.51%	12.00%
MUL RED	86.49%	12.06%
MUL COASTAL	86.33%	9.72%
MUL NIR	86.31%	12.44%
MUL YELLOW	86.16%	9.89%
MUL BLUE	85.45%	12.42%
MUL NIR2	85.17%	11.09%
PSH REDEDGE	80.03%	5.04%
PSH COASTAL	79.93%	2.35%
PSH YELLOW	79.80%	2.37%
PSH NIR2	78.08%	3.16%
PSH BLUE	75.90%	11.26%
Grand Total	91.51%	10.07%

9.3.3 Quality by input spatial resolution

The ranking of the $Q(x)$ obtained by different classes of input image spatial resolution (Table 9.5) shows a top performance of the class C ranging from 1 to 2.5 meter of spatial resolution with $Q(x)$ equal to 92.9%$\pm$8.8%. Class C consists of SPOT-5 and CBERS-2B data and all the multispectral data from the VHR sensors. The second best is class E with spatial resolution ranging from 5 to 10m. and achieving $Q(x)$ = 92.13%$\pm$8.12%. Class E consists of SPOT 1,2 P data and multispectral SPOT-5 and RapidEye data. Interestingly, the worst ranking achieved according to the Q metric, is class A containing all VHR data, i.e. with image resolution of 0.5m. or better. Fine-level of details seems to decrease of the signal/noise ratio that can be observed. The reliability of the GHSL automatic image information extraction workflow under test is maximised for resolutions in the range of 1 to 2.5m and based on this observation future releases of the workflow will be configured with a standard resolution of input imagery to approximately 2.5m. before any feature extraction.

9.3.4 Quality by ecoregions

Apart from the analysis of the GHSL output quality with respect to the data characteristic (sensor, spectral band and spatial resolution), also the landscapes mapped may have an impact on the processing quality. It can be assumed that similar landscapes as for example described by global ecoregions [Ols+01], may introduce similarities in the behaviour of any given automatic image information recognition strat-

Table 9.5 Average quality by class of resolution.

Resolution	Values Average of $Q(x)$	StdDev of $Q(x)$
$C : \{1 < x \leq 2.5\}$	92.90%	8.87%
$E : \{5 < x \leq 10\}$	92.13%	8.12%
$D : \{2.5 < x \leq 5\}$	86.83%	12.59%
$B : \{0.5 < x \leq 1\}$	83.46%	13.27%
$A : \{x \leq 0.5\}$	81.97%	11.27%
Grand Total	91.51%	10.07%

egy. In other words, they can introduce a dominant characteristic of the background of the image, but they can also probably contribute in the explanation of the specific materials and patterns used to make human settlements. Consequently, it is of interest to test the robustness of a specific global automatic image information retrieval task, against a bias introduced by dominant landscape patterns available at a local scale. On the other side, under given constraints it could be possible to extrapolate the performances of a given automatic image information extraction task to images representing the same landscape.

Figure 9.11 shows the average $Q(x)$ obtained by the current I2Q GHSL workflow after extrapolation to the WWF ecoregions[1]. According to this analysis, Brazil, Europe and China are well placed in the high confidence area with $Q(x) \geq 80\%$ together with large parts of the Sub-Saharan and Southern Africa.

Middle East, Sahara and North African areas apparently show systematic problems with the current image information extraction strategy, producing average $Q(x)$ in the range 70% to 80%. One of the probable reasons behind these poor performances is the presence of scattered vegetation and very bright soil background that might create false alarms in the textural image feature and/or miss detection in the morphological image query. Similar issues were already addressed in the same areas by a method applying morphological filtering before the image textural analysis. It demonstrated a drastic increase of the performance of the automatic recognition of built-up structures in arid areas having as background bright soil and scattered vegetation [PG11]. This method was not implemented in the current workflow. The latter is focused more on general-purpose processes of morphological and textural features, without chaining them deductively prior to the learning phase. It is expected however that such type of observations can lead to "regionally-adaptive" image information extraction workflows taking into consideration the local and regional landscape and background conditions for each scene to be processed. This is currently under investigation.

Red regions in Fig. 9.11 correspond to areas with major inconsistencies between the current GHSL outputs and the available reference layers. Such outputs are rejected during the first iteration of the I2Q GHSL workflow and are ignored in the compilation of the final GHSL mosaic. They account for a small fraction of the available scenes, yet they are dealt with great delicacy. The reason for this is that the mean $Q(x)$ computed over such regions and extrapolated to the whole ecoregion might be influenced by a single or just a few images and result in poor performance

[1] URL: `http://assets.worldwildlife.org/publications/15/files/original/official_teow.zip?1349272619`

due to the erratic and possibly arbitrary conditions of the input data. Note that inconsistency with the available global BUref doesn't necessarily imply bad quality output.

Figure 9.12 shows an example of a single image with data leading to a poor performance of the Q metric when generalised to the whole available ecoregion in Bangladesh. The image is a SPOT Panchromatic scene covering the city of Dhaka and its northern-eastern surroundings. The top-left image shows the GHSL at 1:50K output overlapping (50% transparency) with the low-resolution BUref that was used to derive automatically the $Q(x)$ metric. The observed mismatch between the two sets justifies the low $Q(x)$ scores. During the first iteration of the final mosaicing module this output will be discharged, while another dataset coming from a VHR image and covering only the central part of the city will be accepted because of a better matching with the low-resolution reference data (top-right). A visual check of the original image data at a more detailed scale reveals however that the reference information layer in this example provides a strongly biased representation of the actually existing structures, underestimating dramatically the presence of built-up areas in the region. The reason is probably the fact that the settlements, as shown in the bottom-left image, are relatively scattered and contain large vegetated areas. The scenes are discarded by the low-resolution radiometric classification procedure as shown in the bottom-right image.

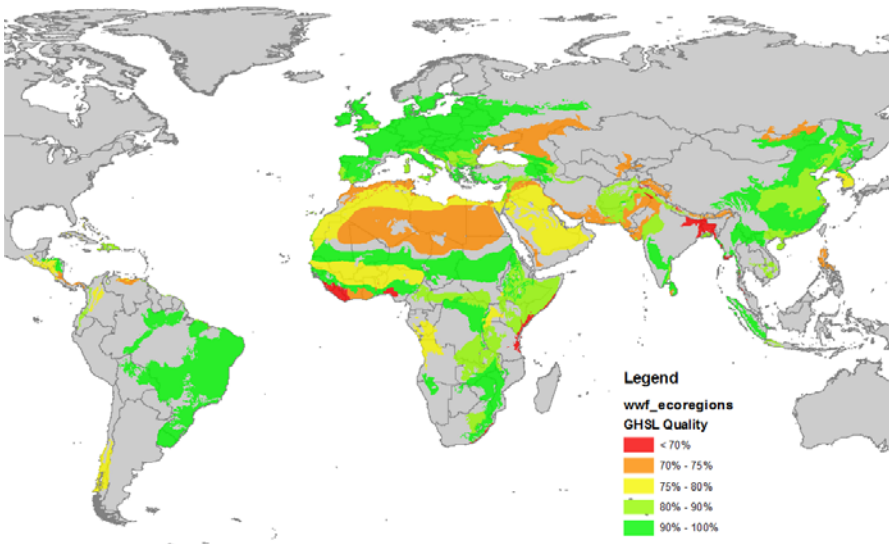


Fig. 9.11 Estimated quality of the current GHSL output by WWF ecoregion.

9.4 Information discovery

A total of 2895 CBERS-2B satellite scenes, accounting for 3.19E+06 km^2 of ground surface, where selected for testing the "discovery" modes of the IQ GHSL learning and classification workflow. For this set of scenes all adaptive learning methods

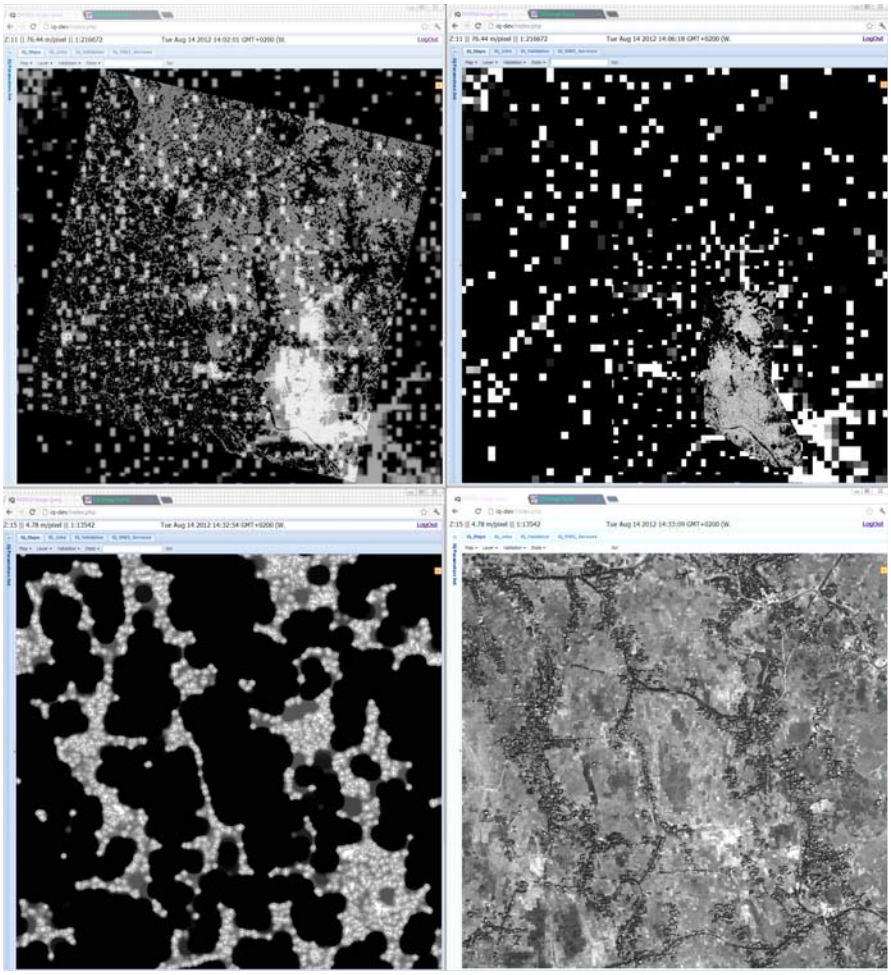


Fig. 9.12 City of Dhaka, Bangladesh. Top-left: available GHSL output information overlaid with the low-resolution reference layer; top-right: final mosaic selecting only those outputs that fit with $Q(x) \geq 85\%$; bottom-left: zoom in the discharged GHSL areas that do not fit sufficiently with the BUref layer; bottom-right: visual check of the input image. The area is densely inhabited but because of the scattered patterns and vegetation it is labelled as not urban in the low-resolution reference layers.

failed due to the absence of information on the presence of built-up areas. Most of these scenes are located in remote rural areas with scattered small scaled settlements away from larger agglomerations (villages and/or towns). In the first iteration of the IQ workflow the BUref layer consisted of a combination of low-resolution globally available layers (LandScan, Modis500), in which low density rural areas are poorly represented. During the 'meta-learning' phase, the statistical distributions of the best thresholds on image features were analysed in the satellite scenes, where the adaptive learning phase was providing high reliability. The satellite scenes chosen for the 'meta-learning' phase were selected by a query listing all the CBERS-2B scenes, where at least 20% of surface was identified as built-up in the available BUref, and producing an output with BUref MER less than 10%. The thresholds

learned from this phase were then applied blindly to the satellite scenes suitable for the 'discovery' phase.

The sum of the new built-up surface discovered in the HR input images during this phase was estimated as 1.00E+05 km^2. This was with an average accuracy of 98.57% and a StdDev of 2.42% with respect to the BUref. The accuracy was measured as the inverse of the MER using BUref as reference.

9.5 Mosaic quality assessment

The input image data models for this study were heterogeneous including i) a collection of individual scenes, ii) a collection of sub-tiled scenes, iii) pan-sharpening fusion of specific bands of individual scenes, and iv) large mosaics made of individual aerial photographs. Problems associated with such image datasets are overlap in the spatial domain and repeated image acquisitions during the time line. Moreover, due to specific sensor characteristics, local landscape and other operational parameters, images of the same area may yield results of rather different quality. For example multispectral acquisitions of the same area may yield a very different output compared to the panchromatic channel used alone, even when using the very same automatic image information extraction method.

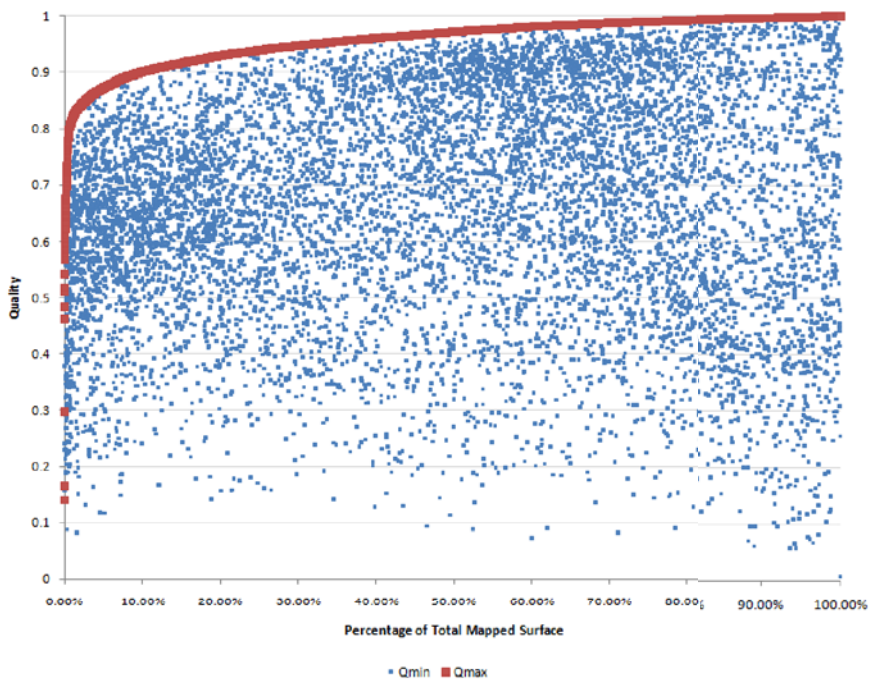


Fig. 9.13 Ranking of all the scenes processed during the experiment by increasing quality optimised among all the processed options (band, learning parameters) available on the same scene. This is the process implemented for the composition of the final GHSL mosaic, taking the best of the available processed pieces of information. The blue dots represent the worst quality available on the same corresponding scenes.

During the mosaic phase this complexity is reduced i) by standardisation of the global projection and scale using the OSGeo TMS standard and ii) by introducing a set of hierarchical composition rules, giving priority in representation of the available information that is judged more suitable of answering to the specific user query. The process is implemented through a recursive loop that is looking for the best information available for each output mosaic tile. Because of the embedded hierarchical composition rules based on optimisation, it is expected that the final GHSL mosaic will have a quality greater than the average of all the available GHSL outputs in all the processed image data. For each area mapped systematically the output providing the smallest estimated error is selected among all the available outputs.

Figure 9.13 shows a simulation of the effect of this mechanism in the universe under processing during the experiment. All the input images processed during the study are ranked by increasing quality optimised among all the output options (band, learning parameters). The best and the worst options available for each input scene are represented in red and in blue respectively. The spread and density of the blue dots below the optimised red line shows the decision space explored by the I2Q system during the composition of the final image information layer. Thanks to this optimisation mechanism, the final mosaic is expected to have a rejection of 10% of the mapped surface, selecting $Q(x) = 90\%$ as the minimal quality threshold.

9.6 Data compression

The data compression is a crucial aspect of any system having the aim of mapping large areas with HR, VHR input images. The control of this aspect becomes mandatory, if the system under test is tasked to provide global outputs, where a non-optimised feature space may introduce an unrealistic explosion of the storage size of the image-derived features and information. Table 9.6 reports about the results of the experiment concerning the image data and information storage.

Table 9.6 Compression ratio of image features and information outputs.

	Data volume (Gb)	compr. ratio 1:x
Source image data	4676	
Image features	1843	2.54
Image information (classification)	366	12.78

Thanks to the fast image feature extraction and classification algorithms applied during the workflow, the I/O throughput is actually becoming the major bottleneck of the current version of the I2Q system: we estimated an average cost of processing (feature extraction and classification) of the imagery to be equal equal or less than 50% of the cost of reading the imagery from the disk. Consequently, the minimisation of the feature space and its compression become an important issue to monitor in order to evaluate the performance of any system delivering HR/VHR global image information. Using the current setup, including 10m-resolution textural (PanTex) and multi-scale morphological features (CSL) a compression ratio of

1:2.54 was observed with respect to the size of the heterogeneous input images used during the experiment. The extracted image information from the available features report about i) the presence of built-up areas and ii) the scale (size) of the included built-up structures. This image-derived information has the same resolution as the image features, but they show a much better compression ratio of 1:12.78, if compared with the original size of the image data.

Chapter 10
Case study: feasibility of European GHSL

10.1 Introduction

The Joint Research Centre, in pursuing its mission of scientific and technical support for policy making in Brussels, has planned and carried out a series of tests in collaboration with DG Regional Policy, in view of a possible integration of its in-house Global Human Settlement Layer (GHSL) technology to the European Urban Atlas (UA). The support is based on specific parameters derived from the application of the satellite-based methodology developed by the JRC for human settlement analysis. The Urban Atlas provides detailed and cost-effective digital mapping, ensuring that city planners have the most up-to-date and accurate data available, offering new tools to assess risks and opportunities, ranging from threat of natural disasters and impact of climate change, to identifying new infrastructure and public transport needs. The GHSL/UA integration would contribute to population disaggregation and risk and disaster management applications, as well as support regional planning in general.

This feasibility report describes the application of the GHSL protocol according to the Urban Atlas product specifications and more specifically the comparison between European Soil Sealing Layer (SSL) output information with the GHSL built-up information extraction in the context of the Urban Atlas 2012-2013.

The objectives of the work described in this report are i) to test the processing capacity of the JRC IQ system in order to assess the feasibility of a pan-European GHSL coverage or ”built-up-areas detection” using the image data prepared for the UA 2012-2013, ii) to assess the reliability and added value of the automatic image information retrieval by systematic comparison of the automatic output with a known reference layer reporting about similar information, namely, the European soil sealing layer (SSL).

10.2 Input data

10.2.1 Image data

The image data potentially available for the current test is made of 793 satellite scenes collected by the SPOT-5 platform during the years 2003-2009. The data was

made available by the European COPERNICUS Programme (previously known as Global Monitoring for Environment and Security) for the establishment of a European capacity for Earth Observation. Originally it was acquired to produce the Urban Atlas 2006.

Count of imageid	Column Labels			
Row Labels	MUL	PAN	PSH	Grand Total
4326			23	23
32628			2	2
32629	13	12	26	51
32630	13	7	122	142
32631	4	8	63	75
32632	10	23	136	169
32633		19	126	145
32634	13	9	124	146
32635	5	6	29	40
Grand Total	58	84	651	793

Fig. 10.1 Number of scenes by scene projection and type of image data.

less than 0.1%	260
from 0.1% to 10%	331
from 10% to 20%	18
from 20% to 30%	8
from 20% to 40%	4
from 40% to 50%	2
more than 50%	1

Fig. 10.2 Number of scenes by Cloud Cover percentage.

The satellite scenes are projected in various local UTM projections and also geographical lat-lon projection (EPSG 4326). These scenes have different radiometric characteristics including multispectral, panchromatic and "pan-sharpened" products (Fig. 10.1). Moreover, these input scenes show various levels of cloud cover. 260 scenes show cloud coverage less than 0.1%, while all the others show cloud coverage up to 50% (Fig. 10.2). The expected spatial displacement or tolerance admitted is of 5m RMSE, as specified in the UA technical specifications.

Figure 10.3 shows the spatial distribution of the available image data and the sensor characteristics. We can observe that while the majority of areas are covered by PSH scenes, in some areas only PAN image data are available, while occasionally some PAN+ MUL bundle images are also available. In order to simplify the design of the test, only PSH images having a UTM projection were introduced in the processing list. In total they are 628 scenes scattered all around the European territory (Fig. 10.4).

With respect to the standard IQ GHSL workflow the available image data for this test is placed between the 1:10K and 1:50K GHSL specifications. In particular, the geo-coding quality would allow 1:10K GHSL products, but the input image resolution is not enough to calculate the morphological/shape criteria that are necessary to recognise and describe the built-up structures at 1:10K scale. Moreover, the pan-sharpening adopted for the fusion of multispectral and panchromatic data, although producing suitable results for visual inspection and interpretation, has some draw-

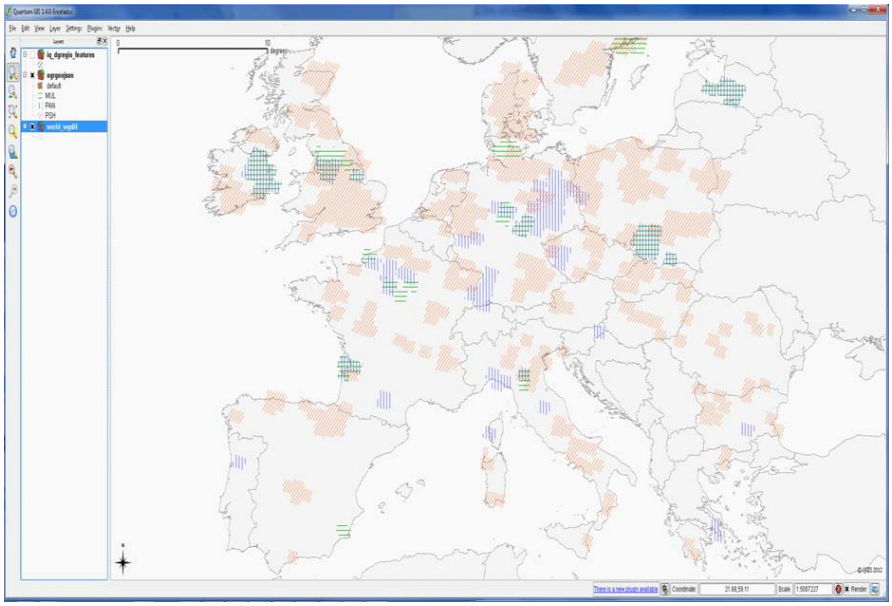


Fig. 10.3 Available image data input by type of image data.

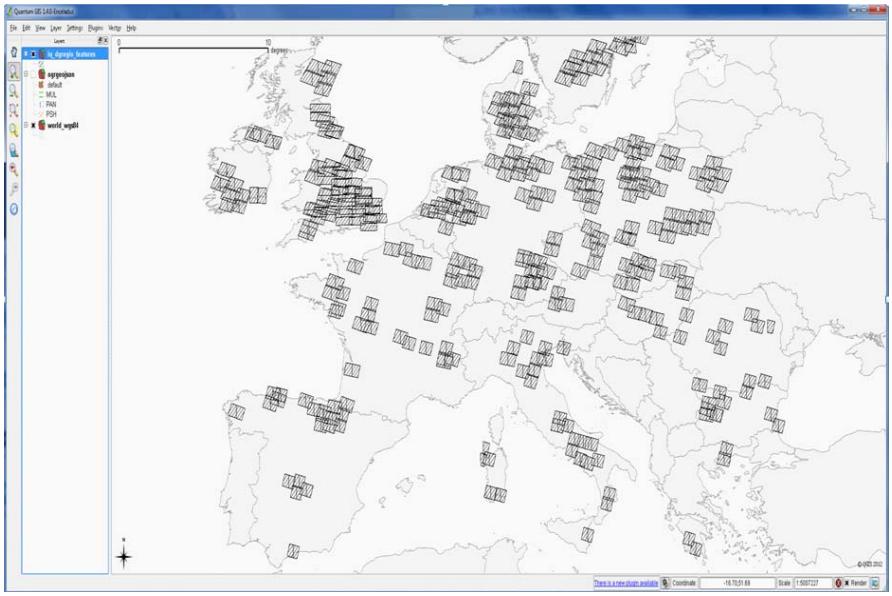


Fig. 10.4 Image data input under test.

backs on the side of the automatic exploitation of the same input image data. The image pan-sharpening degrades the image radiometric criteria, which impact significantly the performances of automatic shadow and vegetation analysis. According to evidences collected in the available images, the pre-processing included also a 'visual enhancement', probably through high-band-pass filtering by convolution. This kind of filtering introduces image artifacts on the borders of the image structures (objects) and increases the overall image noise, decreased the signal/noise ratio. All these facts have a negative impact on the performances of the textural and morphological/shape image-derived criteria needed for the automatic discrimination and characterisation of the built-up structures.

10.2.2 Reference data

In order to maximise consistency of the output, the IQ GHSL image information workflow includes automatic optimisation of some processing parameters by systematic comparison with known reference information sources. This includes two standard reference layers, the global population density layer (LANDSCAN, 2010) with approx. 1 km resolution and a global land use map of urban extents derived from remote sensing data (MODIS500)with 500m nominal resolution.

For this case study we added an additional layer in the workflow that is available only for European countries: Degree of soil sealing with a resolution of 100m, which was produced by the GMES Fast Track Service Precursor on Land Monitoring. This layer is used for the production of the UA urban fabric classes discriminating them in different soil sealing percentage thresholds. It was produced by a hybrid procedure intersecting satellite-image-derived land cover information with data reporting on roads and settlements. It is used here for benchmarking the GHSL product accuracy in detection of built-up areas.

10.3 Workflow

During the current experiment, a mix of the standard 1:10K and 1:50K IQ GHSL queries was implemented in order to cope with the intermediate scale (between 1:10K and 1:50K) of the available input images with respect to the GHSL specifications and data requirements. The output of the two queries is then integrated (by sum) and evaluated using the reference sealed soil surface layer at 100 meters of resolution (Fig. 10.5).

In particular, the recognition of built-up areas are done at 1:50K scale, while a subset has been used, if the 1:10K GHSL workflow was applied for detection and characterisation of single built-up structures inside these detected built-up areas. The characterisation was based only on the estimated area (in plant) of the built-up structures. At 1:50K scale, any spatial reference unit (pixel) of 50x50 meters that covers at least one built-up structure or part of is considered built-up. Analogously, at 1:10K scale any spatial unit of 10x10 meters, covering at least one building or a part of it is considered built-up. At 2.5-m input spatial resolution, only 1:50K built-up areas are reliable: moreover, the tolerance in characterisation of the area of built-up structures is of 6.25 meters (surface of one input image pixel).

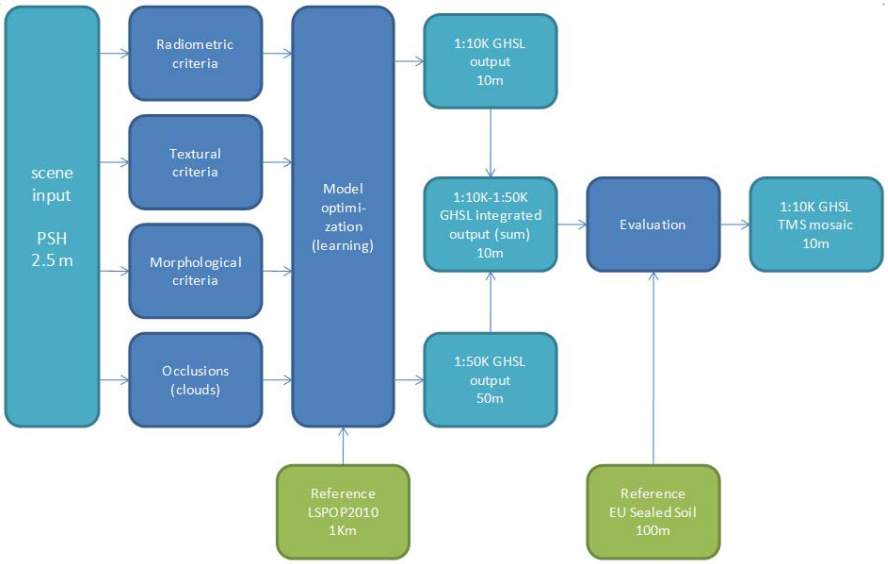


Fig. 10.5 Automatic image information workflow applied during the experiment.

The two GHSL output scales are then integrated by sum at the 1:10K scale resolution: in this way we create a hybrid 1:10K-1:50K product that is evaluated by systematic comparison with the European SSL.

10.4 Evaluation parameters

Both the reference layer and the GHSL output are continuous information layers: the reference layer reports about the percentage of sealed surfaces in respect to the spatial unit of 100x100 meters, while the GHSL output reports about the presence of built-up structures in the spatial units of 10x10 and 50x50 meters, integrated by sum.

The built-up presence information released by the IQ system at any scale is a membership value to the class built-up as formalised by the image information query. For the purpose of this report, this membership value can be interpreted both i) as probability (or possibility) score made by the system that this cell corresponds to a built-up structure on the ground, and ii) as a percentage of the unit cell corresponding to a built-up structure. The two statements can be merged in one by interpreting the membership value as the estimated possibility that the whole output unit cell is covering a built-up structure on the ground.

In order to measure the agreement between the GHS output and the reference, we adopted two strategies: i) directly measure the agreement of the continuous information layers by linear regression techniques, and ii) dichotomise the continuous information layers by a given threshold, and then calculate agreement measurements based on confusion matrix. In this experiment, a threshold of 25% in both layers was considered to discriminate between built-up (BU) and not-built-up (NBU) classes. In the dichotomic-classification case the GHSL output under test was aggregated

at the same resolution provided with the reference layer, then 100x100 meters. In the continuous case two different resolutions were put under test: 50x50 meters and 500x500 meters.

During the evaluation, the following measurements were collected per each input scene processed: `overall accuracy`, `built-up accuracy`, `built-up agreement`, and `x_fit`.

- The `overall accuracy` is the number of pixels with agreement on the BU/NBU classification divided by the sum of all the pixels analysed in the scene;
- The `built-up accuracy` is the number of pixels with agreement on the BU class divided by the sum of all BU reference pixels analysed in the scenes;
- The `built-up agreement` A is a per scene global measurement expressing the agreement on the total surface classified as BU class:

  $$A = 1 - abs(\texttt{pixBUclass} - \texttt{pixBUref})/(\texttt{pixBUclass} + \texttt{pixBUref}),$$

  with `pixBUclass` be the number of BU pixels estimated by the classification under test and the `pixBUref` the number of BU pixels estimated by the reference layer;
- The `x_fit` measurement reports about the per-pixel R-square linear regression fit (correlation) between the GHSL output and the reference layer. It was calculated using two different scales or spatial resolutions: 50x50 meters, and 500x500 meters.

10.5 Results

10.5.1 Qualitative inspection

Figure 10.8 shows the standard output obtained by the GHSL workflow under test over the city of Luneburg, Germany. Brighter grey levels mean higher membership output value to the class ”built-up” which is automatically generated by the IQ system. Figure 10.6 shows the same region as represented by GoogleEarth imagery, while Fig. 10.7 shows the representation made by the SSL reference layer. Also in this case brighter grey level means higher value of ”sealed surface” percentage.

Figure 10.9 shows the GHSL output concerning the variable: ”size of built-up structures”. For ”size” here it is understood the estimated area (in plant) of the built-up structure. The original GHSL information is represented in square meters or ”scale”: for visualisation purposes the colour palette applied ranges from green (low) values to red (high) values.

10.5.2 Quantitative analysis

In this test, only the built-up area detection of the GHSL output was evaluated quantitatively. Other GHSL output information layers including characterisation of built-up structures inside the built-up areas (discrimination by size) were not taken in to account. The reason of this choice is twofold:

Fig. 10.6 The city of Luneburg, Germany at 10-m-resolution.

1. built-up areas detection is preceding all other GHSL output measurements in the workflow, including built-up areas characterisation. Hence it is the most important variable influencing the reliability of all the others;
2. the impossibility to implement a consistent quantitative evaluation protocol — without having a consistent reference layer of a suitable scale— as a single building footprints at 1:10K scale.

The results of this first quantitative test show a good level of agreement between the built-up areas automatically generated by the JRC IQ system following a customised GHSL task and the reference SSL layer.

The agreement is good in both dichotomic and continuous evaluation schemata (Figure 10.10). In particular, the 628 satellite scenes under test were showing a

Fig. 10.7 The same region as represented by the SSL reference layer.

90.8 ± 3.9 average *overall accuracy* rate and 86.6 ± 7.1 average rate of *built-up accuracy*. Moreover, the correlation fit is estimated as 84.7 ± 8.0 and 96.1 ± 5.1 at 50m and 500m of spatial generalisation.

Because of the different nature of the two compared information layers, these measurements include both errors in the classical sense as well as errors due to the different semantic and scale/generalisation parameters. In other words, the 10% of disagreement between the two information layers can be originated by the following phenomena:

1. Wrong built-up area value in the GHSL output if compared with reality (input imagery);

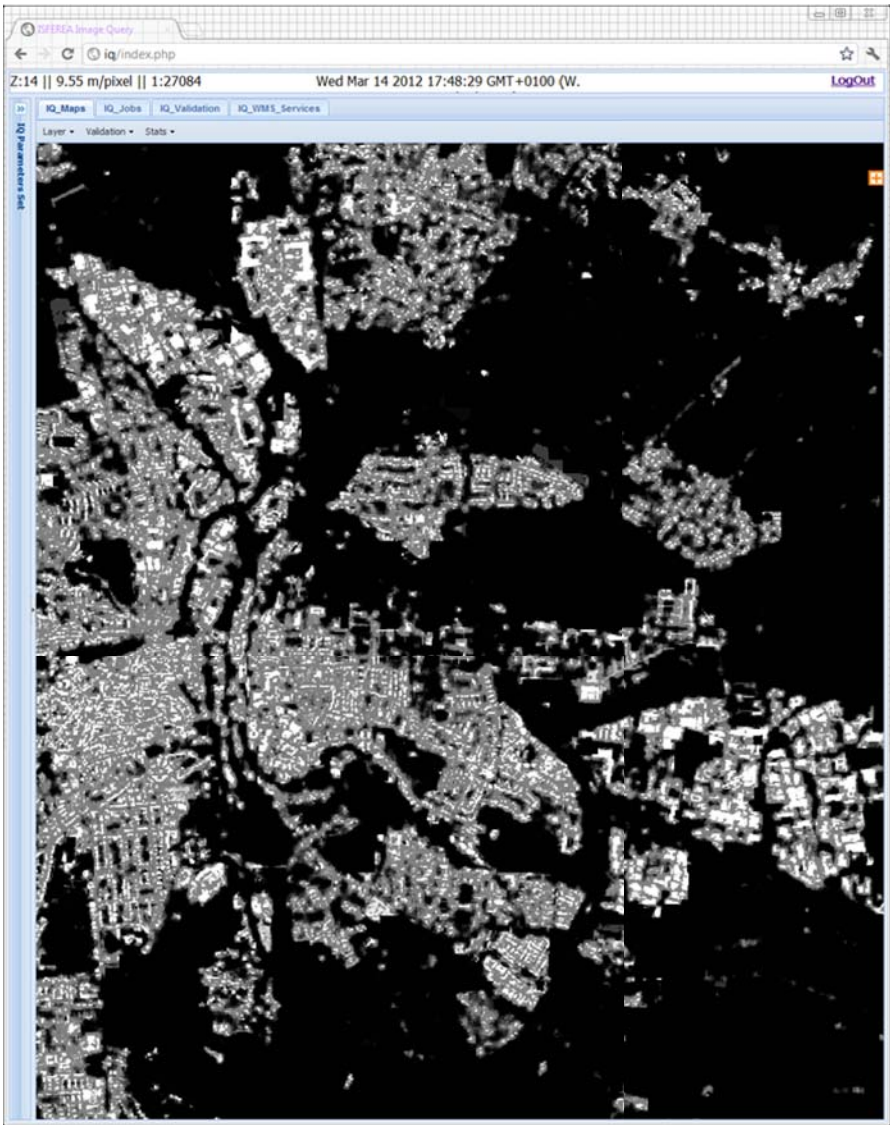


Fig. 10.8 The same region as represented by the ”built-up areas” GHSL output.

2. Wrong sealed percentage value in the reference layer if compared with reality (input imagery);
3. Different scale and generalisation parameters;
4. Different semantic definition of built-up area and sealed surface area.

While the first two items are generally considered being error scores, the last two are more related to relative differences in the definition of the geographical information in the two layers under comparison. The discrimination between these different phenomena is quite challenging and may require expensive additional independent assessments following the two distinct geographic information collection protocols.

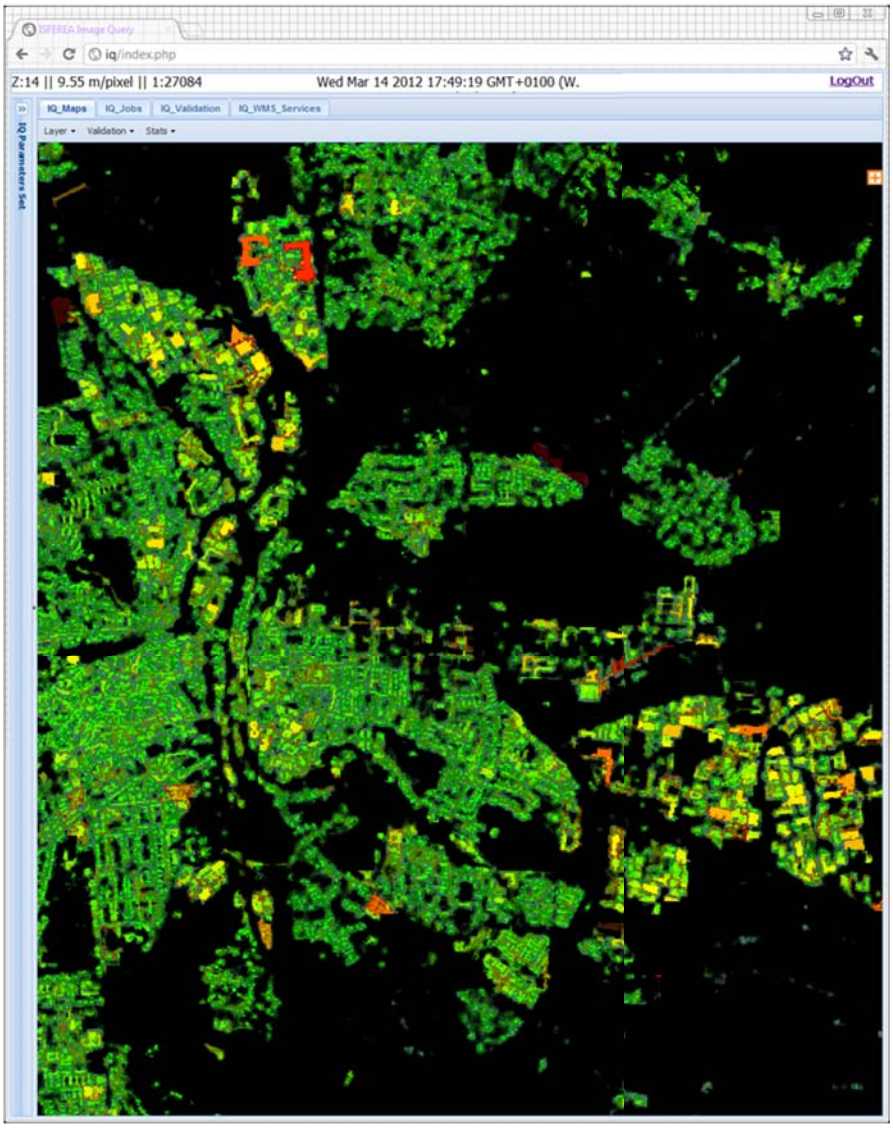


Fig. 10.9 The same region as represented by the ”size of built-up areas” GHSL output.

	Overall accuracy	BU_accuracy	BU_agreement	x_fit(50)	x_fit(500)
mean	90.82%	86.64%	87.46%	84.75%	96.17%
stdv	3.99%	7.10%	12.49%	8.09%	5.11%

Fig. 10.10 Agreement measures between the GHSL *built-up areas* output and the reference layer.

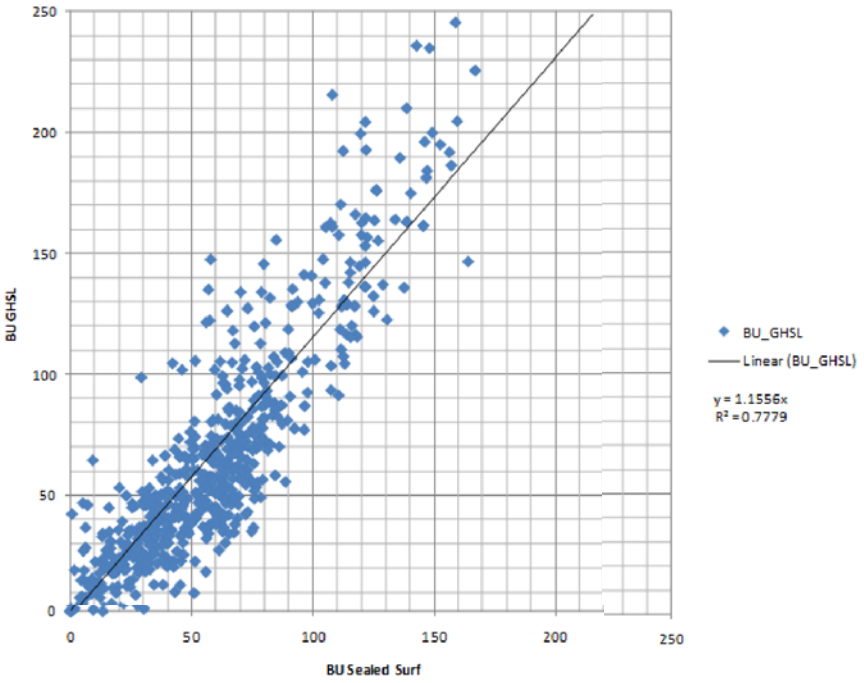


Fig. 10.11 Correlation between the total estimation of BU surface per scene in the reference and in the GHSL layers: all satellite scenes under test.

Certain attempts can be made by observing that the factors related to the points iii) and iv) have more chance to take the behaviour of as a systematic bias effect, while factors i) and ii) are in principle randomly distributed .

Figure 10.11 shows the correlation between the total (per scene) surface estimated as ”built-up” by the SSL and the GHSL layers, for all the scenes under test. It can be noticed a slight positive slope in the estimated linear regression (y = 1.15*x) meaning that on the whole set under study the GHSL estimation of BU areas tends to be systematically more abundant than the one made by the SSL layer.

This is a counter-intuitive result, because in principle ”built-up” surfaces are expected to be a sub-set of the ”sealed” surfaces. Hence, it would be expected to find systematically more sealed surface than built-up. The expected slope factor in the linear regression would be smaller than 1. This result can be partially explained by observing the linear regression in the subset of scenes having low density settlement pattern.

For example, in Figure 10.12 the regression is estimated only in the subset of scenes having less than 100 square kilometres of built-up surface, and excluding all the scenes dominated by big urban centres. In this case, the linear regression takes the expected slope smaller than 1 (y = 0.95*x) and showing a systematic underestimation of GHSL ”built-up” surfaces with respect to the ”sealed” surface.

Even if requiring further analysis, we think that these empirical evidences are the combined effect of the two ”disagreement” factors described above: the iii) different scale and generalisation parameters and the iv) different semantic definition of ”built-up area” and ”sealed surface area”.

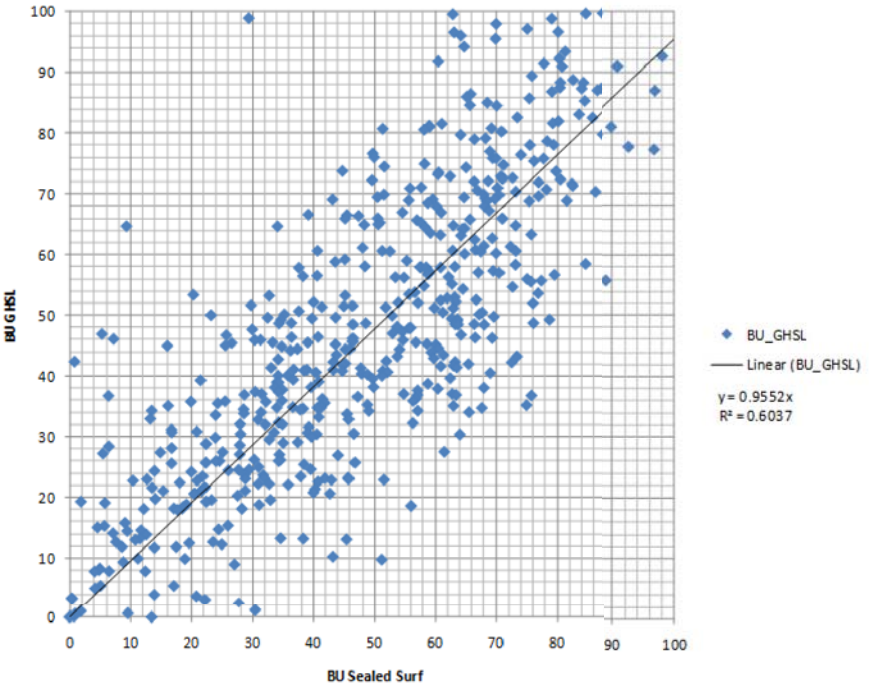


Fig. 10.12 Correlation between the total estimation of BU surface per scene in the reference and in the GHSL layers: it is considered only the subset of satellite scenes having less than 100 square kilometres of BU surface.

The working hypothesis is that, while the factor (iv) is always related to the background, the factor (iii) becomes dominant in case of presence of compact large urban nuclei, where the surface of open spaces and roads between buildings is highly relevant respect to the total area under analysis. This ”interstitial” surface is then changing BU/NBU class according to the different scale and generalisation rules adopted in the two layers under comparison. As known, the impact of scale and generalisation factors is heavily related to the spatial pattern of the represented information.

10.5.3 Difference maps

During the test ”difference maps” were systematically calculated for all the scenes under analysis in order to help the identification and understanding of the agreement/disagreement of spatial patterns between the two information layers. In this first test, only dichotomic difference map where calculated using the same spatial resolution and thresholds used for the accuracy analysis discussed above.

Figure 10.16 shows an example of the ”difference map” obtained in region southeast of Bologna Italy. Red and blue pixels represent areas where the SSL overestimates and underestimates, respectively, the GHSL built-up areas. Figure 10.13 shows the same region from GoogleEarth imagery for visual inspection, while Fig-

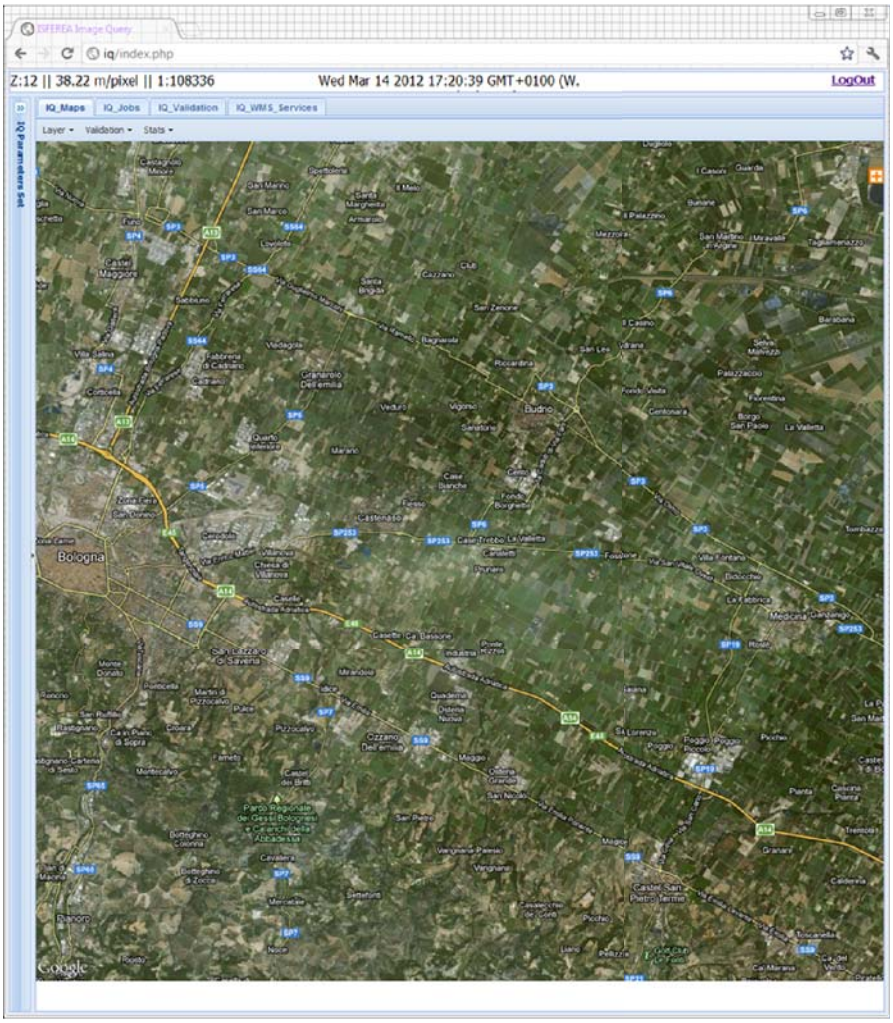


Fig. 10.13 Example region South-East of Bologna, Italy.

ure 10.14 and Figure 10.15 represent, respectively, the SSL and the GHSL information layers.

Some patterns observable from Figure 10.16 are well explained by the different semantics embedded in the two products: for example the highway is present in the SSL layer while it is not reported by the GHSL output, because it is not matching with the ”built-up” definition.

Other patterns (usually large red patches, relatively isolated) of sealed surface overestimation with respect to the GHSL output can be explained instead as errors of the SSL detection. See for example Figure 10.17 showing two ”sealed surface” patches in reality corresponding to agricultural fields.

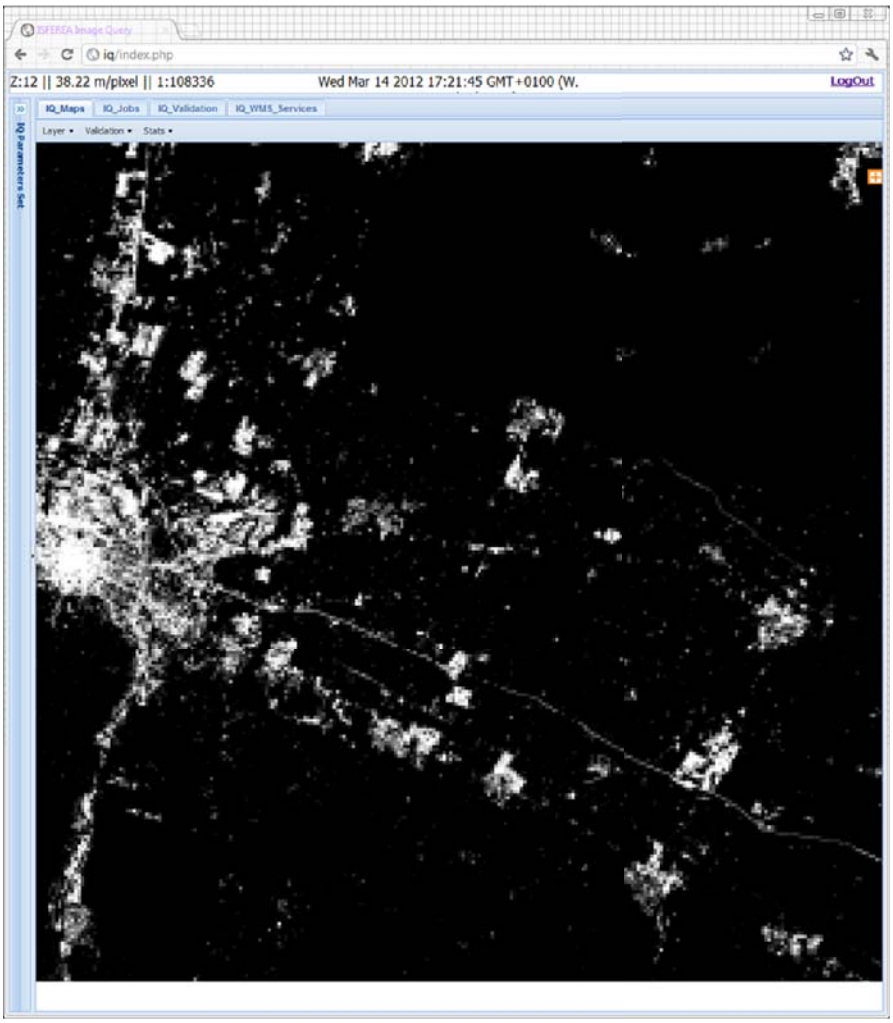


Fig. 10.14 The same region as represented by the SSL layer.

10.6 Conclusions

10.6.1 Assessment

A test of the JRC GHSL production workflow and output was done by JRC during February 2012 with the purpose of:

- testing the processing capacity of the JRC IQ system in order to assess the feasibility of a whole European GHSL coverage of ”built-up areas detection” using the image data prepared for the UA 2012-2013;
- assessing the reliability and added value of the automatic image information retrieval by systematic comparison of the automatic output with a known reference layer reporting similar information, namely, the European sealed soil surface layer.

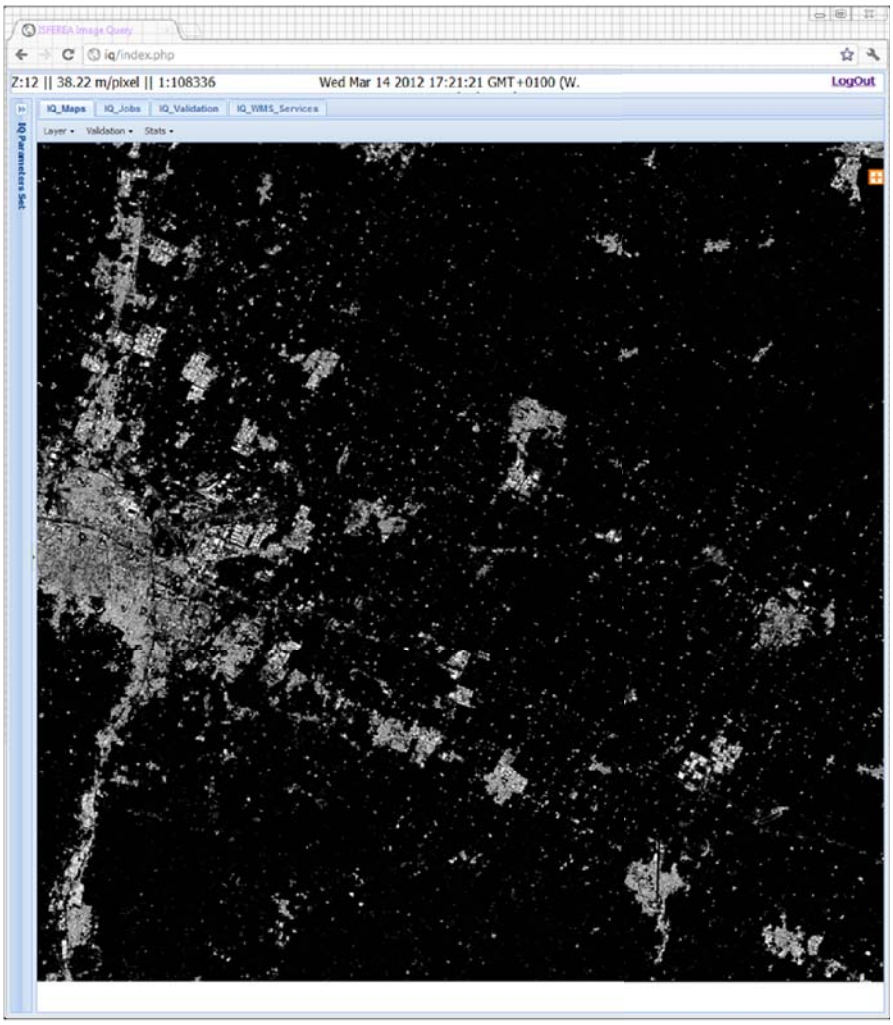


Fig. 10.15 The same region as represented by the ”built-up areas” GHSL output.

During the test 628 satellite scenes used for the production of the UA were successfully processed by the JRC IQ system, performing automatically a customised version of the standard GHSL query including i) automatic recognition of built-up areas and ii) automatic characterisation of built-up structures based on size (area). Automatic detection of occlusions and no-image-data areas (clouds) and hierarchical mosaicing were also activated in order to test the production of seamless information layers from any arbitrary set of partially-overlapping or cloud-covered input satellite scenes. Other standard GHSL descriptors (building height, vegetation) were not calculated because of technical limitations of the input images available.

The computation and production test was successful and showed the capacity to produce seamless European layers using the same JRC technology and similar input data. Limiting factors are substantially linked only to the available storage space capacity of the current IQ system that should be up-scaled in order to i) improve the spatial resolution (the scale) of the output information layer now set to 10 me-

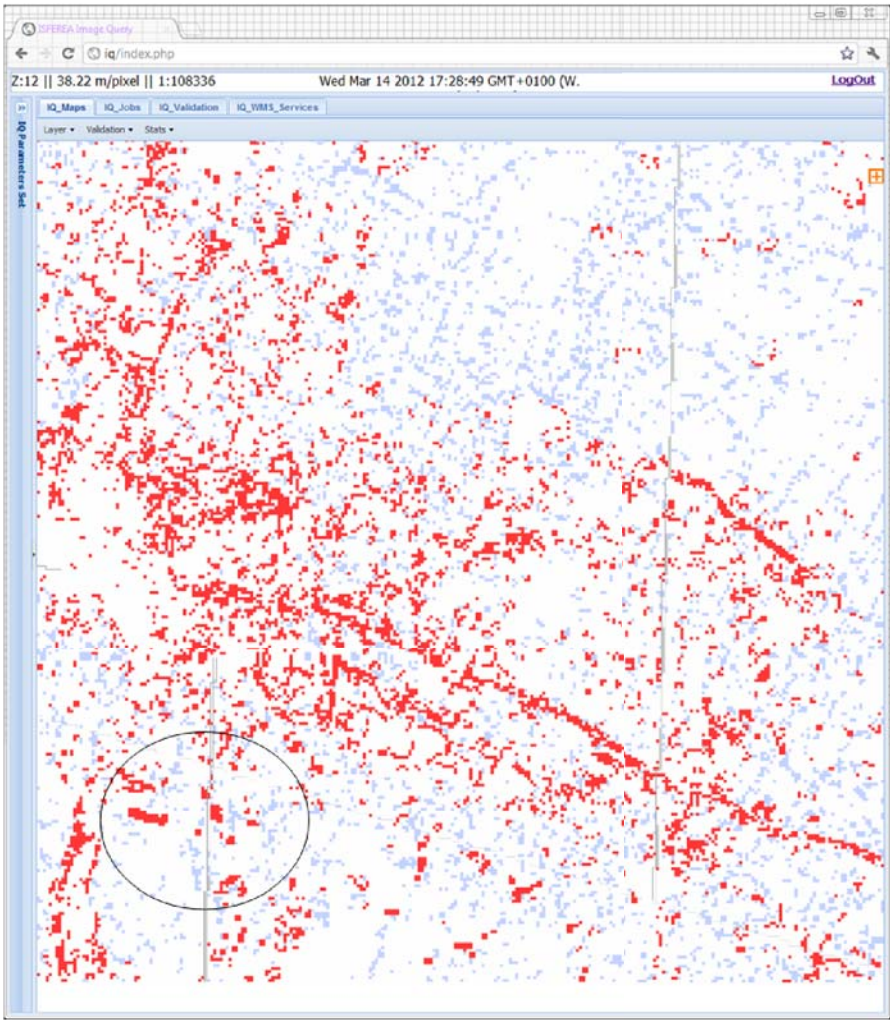


Fig. 10.16 Difference map between the dichotomic "built-up areas" and "sealed surface" classes. In red and light blue the positive, negative differences, respectively, in the Sealed/GHSL comparison.

ters ii) improve the number of information layers that can be produced (settlement characterisation)

The results of the automatic recognition of built-up areas made by the IQ system were systematically compared in all the 628 processed satellite scenes with the SSL layer used for the production of the UA. According to the results of this comparison, the IQ GHSL automatic output showed an overall high degree of agreement in respect to the SSL layer. The agreement is good in both dichotomic and continuous evaluation schemata. In particular, the 628 satellite scenes under test were showing a 90.8 ± 3.9 average "overall accuracy" rate and 86.6 ± 7.1 average rate of "built-up accuracy". Moreover, the correlation fit is estimated as 84.7 ± 8.0 and 96.1 ± 5.1 at 50 and 500 meters of spatial generalisation, respectively.

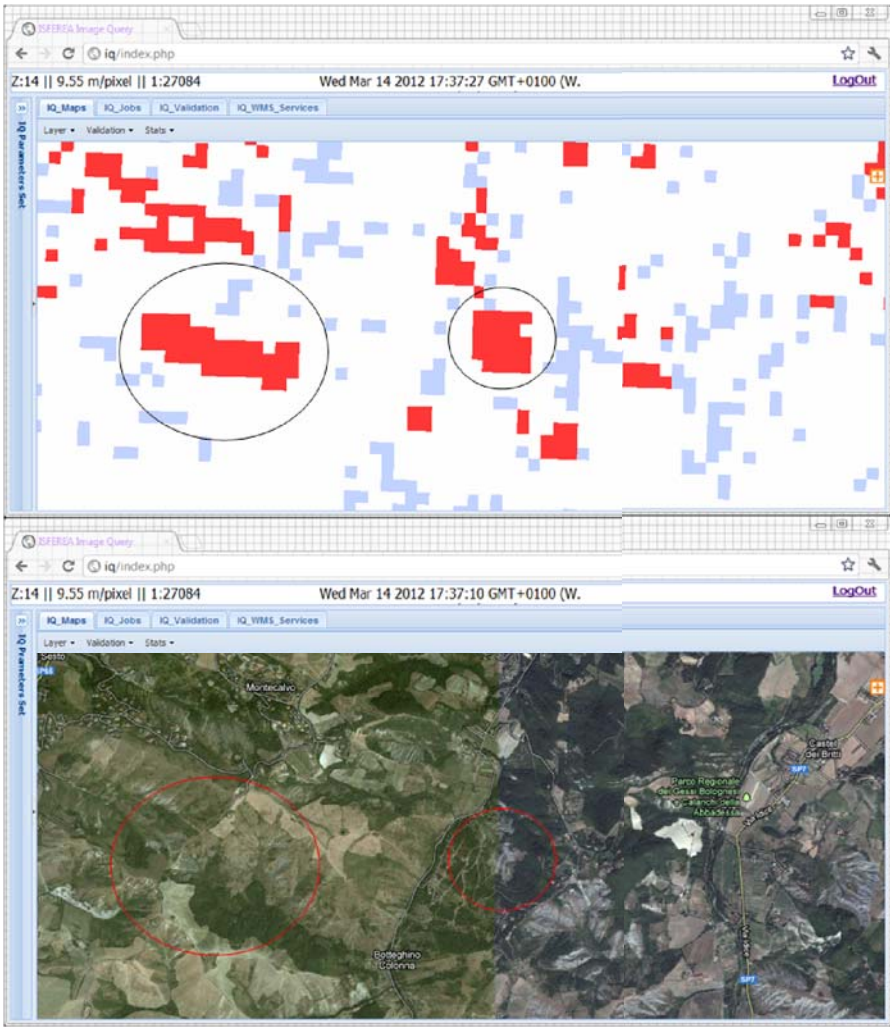


Fig. 10.17 Zoom in the difference map revealing some error spots in the SSL reference layer.

10.6.2 Next steps

The analysis of the 10-20% thematic differences between the two layers has only started. We can observe the contribution of four main factors: i) commission errors in GHSL; ii) commission errors in SSL; iii) different scale and generalisation parameters; and iv) different semantic definition of ”built-up area” and ”sealed-surface area”.

Better understanding of the contribution of these factors would require the design of an improved experimental setting including:

1. Independent collection of reference information from the same input images by an explicit protocol including visual interpretation rules and a clear definition of ”built-up” areas;

2. Increased resolution of the GHSL output under test in order to better describe the settlement components including built-up structure morphological characteristics and the space between built-up structures (road, open spaces);
3. Production of increased number of descriptors of settlement patterns including different levels of generalisation and different constraints in the available GHSL information layers (morphological characteristics, patterns).

Chapter 11
Case study: feasibility of Brazilian GHSL

11.1 Introduction

This chapter describes an assessment of the feasibility of automatically mapping the settlements of Brazil using the using the data of the High Resolution Panchromatic Camera (HRC) operated on board of the China-Brazil Earth Resources Satellite program (CBERS) 2B satellite and the processing workflow of the GHSL described in previous chapters.

11.2 Data

11.2.1 CBERS-2B data

We used 5620 single satellite scenes acquired between 2007 and 2010 by the HRC instrument on board the CBERS-2B satellite. The panchromatic data has a spatial resolution of 2.5m. The scenes were downloaded from the online INPE catalogue[1] allowing for a maximum cloud cover of 50 percent. The scenes cover 1.4 million km2 of the most populated areas of Brazil along the southern and eastern coast as well as areas around Brasilia, Manaus and Cuiaba (Figure 11.1). According to the LandScan[2] population data set this area is inhabited by 118 million people, approximately 60 percent of the population of Brazil.

11.2.2 Ancillary data

Several additional data sets were used in the workflow as ancillary data for the processing or for the validation of the processing output. For the orthorectification of the CBERS-2B HRC data we used the TerraColor as reference layer. This is an orthorectified global imagery base map at 15m spatial resolution built primarily

[1] URL: `http://www.dgi.inpe.br/CDSR/`

[2] copyright by UT-Battelle, LLC, operator of Oak Ridge National Laboratory under Contract No. DE-AC05-00OR22725 with the United States Department of Energy

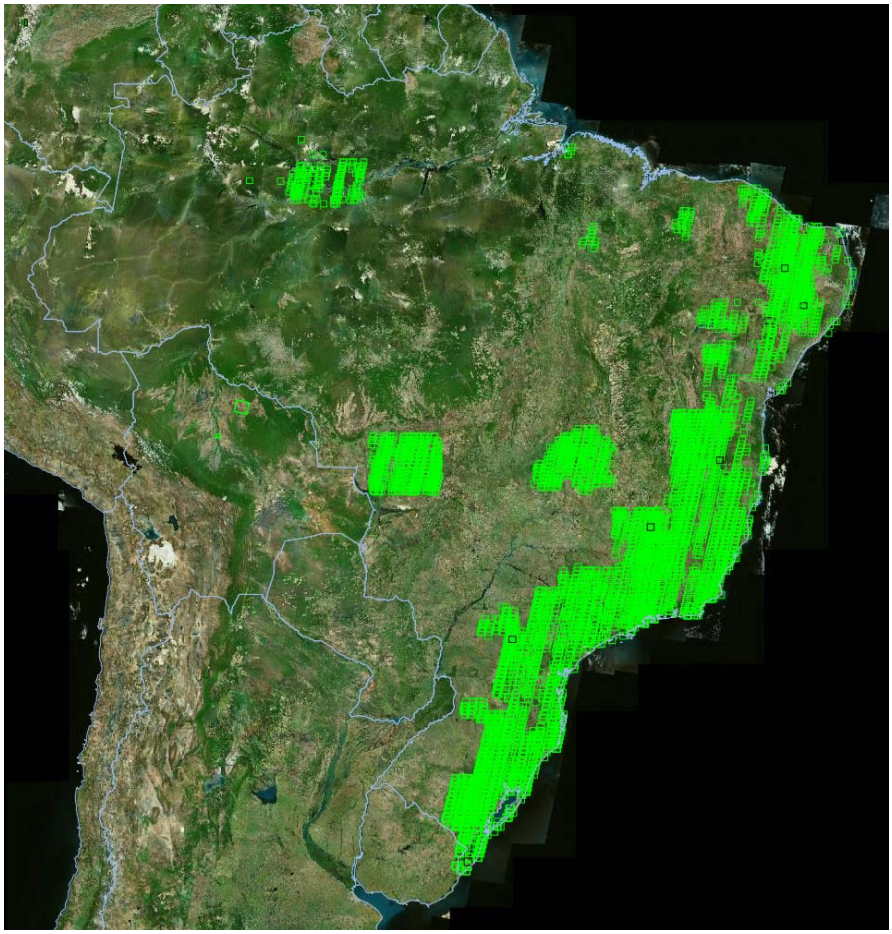


Fig. 11.1 CBERS-2B HRC image footprints used in this study superimposed on TerraColor mosaic.

from Landsat-7 satellite imagery. The Open Street Map (OSM) data were used to extract a high resolution land-sea border. During the processing global data sets are used to control the information extraction output. One of the data sets is urban class of the MODIS Land Cover Type product [SFP10]. In addition the LandScan (2008 and 2010) global population data sets were used. The selection of scenes for the validation uses the WWF ecoregions of the world[3] and the 2010 census of Brazil[4].

11.3 Workflow

In this case study the same the workflow follows the one described in Sec. 4.3. During the pre-processing, the CBERS data are first checked for their spatial consistency. For the CBERS-2 HRC imagery, the geolocation of the input imagery is unre-

[3] URL: `http://www.worldwildlife.org/science/data/terreco.cfm`

[4] URL: `http://www.ibge.gov.br/english/`

liable with the scene centre displaced by up to 40 kilometres [dCE04]. We have used normalised cross-correlation measurements [BS72] to automatically find tie points for correcting the geolocation of the input CBERS imagery. The 15-m-spatial resolution TerraColor Landsat data set is used as a reference to reach a RMS spatial tolerance of around 20 meters, which is sufficient for a mapping of settlements at a map scale of 1:50,000. The CBERS image is then warped using the affine transformation model with least square estimation of the transformation parameters given the final set of tie points. The radiometric consistency check identifies occlusions and no-data areas including clouds. The image information query starts with a feature extraction using both textural and morphological features. The textural image features used in this study are derived from grey-level co-occurrence matrix (GLCM) contrast textural measurements [RKI73]. The contrast textural measures calculated using anisotropic displacement vectors are combined in a rotation-invariant image feature, called PanTex [PGK08], [Pes+11] by using extrema operators. The morphological features are derived from a multi-scale morphological analysis protocol referred to as the "mtDAP" [OPS12]. It computes the Differential Attribute Profile (DAP) vector field [Dal+10] from the input imagery. The DAP of a pixel is the concatenation of two response vectors. The first registers intensity differences, i.e. contrast, within a top-hat scale-scape of an anti-extensive attribute filter. The second registers intensity differences on the bottom-hat scale space of an extensive attribute filter. The mtDAP can be configured with any morphological attribute filter, but in this case simple area openings and closings prove to be sufficient. The area attribute is used to order objects based on size and it is computed incrementally. The protocol organises the image information content into meaningful structures using a Max-Tree structure [SOG98]. The feature extraction is followed by learning procedures in order to optimise the decision thresholds in the input features using a given reference layer. The learning is necessary to process multiple-scene image input data without manual intervention in the input parameter set. After the final classification of the single scenes the quality is assessed using the BUREF layer. This layer is a reclassification of LandScan data set using MODIS as training set. The scenes with the best quality are eventually used for the mosaicing of the final product. In addition to the automated quality control an independent validation is based on a dichotomic validation protocol [Ehr+12] [link to the relevant section of this report]. This includes the collection of spatial samples by a systematic grid procedure and the interpretation of each sample by visual inspection of the corresponding part of image. For each sampling grid a cell of 200x200 meters is selected. Each sample cell is subdivided in 4x4 sub-samples with the size of 50x50 meters and screened for the presence of buildings. For this feasibility study we have used a stratified sampling approach based on the ecoregions of the world and the urbanisation status of Brazil (rural/urban area) derived from the 2010 census of Brazil. The stratified sampling should assure that the validated scenes cover a wide range of ecological zones and building densities. The 98 scenes have been selected in such a way that they represent 50% rural and 50% urban areas (according to the census enumeration areas). In addition the cover the following ecoregions:

- Tropical and subtropical moist broadleaf forests (60%);
- Tropical and subtropical grasslands, savannas, and shrublands (30%);
- Flooded grasslands and savannas (8%);
- Tropical and subtropical dry broadleaf forests (1%);
- Deserts and xeric shrublands (1%).

11.4 Results

The pre-processing of the CBERS2-B HRC data was not trivial due to the poor geolocation accuracy and the potentially high cloud coverage. Out of the 5620 scenes downloaded, we were able to process 3314 (60%). For this study we have applied a learning technique based Receiver Operational Characteristics (ROC) [Kul68] optimisation using MODIS urban areas as a reference. Compared to the 98 scenes that were analysed visually, this is providing an equal error rate (EER) of 0.18 ± 0.09 and a minimal error rate (MER) of 0.06 ± 0.06. The EER/MER is higher in rural areas (0.2008/0.0135) than in urban districts (0.1738/0.1035). With respect to the terrestrial ecoregions, the biggest error (0.233/0.015) is obtained for the xeric shrublands due to confusion with rock outcrops or single trees and shrubs. Figure 11.2 shows an example of the information extracted from the CBERS-2B HRC data for the city of Sao Paolo. The figure shows the presence of buildings, where the pixel brightness is proportional to the percentage of built-up presence in the specific spatial units. For Belo Horizonte several scenes were available and the final result for each spatial unit is made of the 'best' information extracted from all the available processed scenes. Visually, the map shows a high level of detail not only in the urban areas, but also in the rural settlements around the city. Figure 11.3 shows the same information for the Sao Paolo, Rio de Janeiro and Belo Horizonte triangle. Moreover, the high spatial resolution data input offers the option of characterising the built-up areas by automatic analysis of the morphological characteristics of the built-up structures. Figure 11.4 shows a zoom into the area south of the airport of Belo Horizonte. The image shows the 'average building size' layer at a 1:10,000 map scale. The colour-coding follows the blue-green-yellow-red order on increasing size of the built up structures.

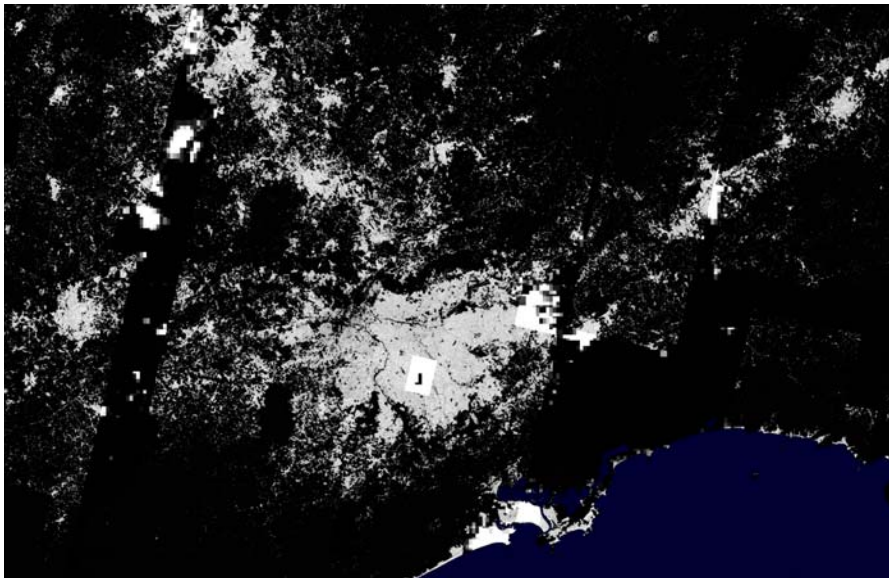


Fig. 11.2 Sao Paolo (Brazil). Presence of buildings with the pixel brightness proportional to the percentage of built-up presence in the specific spatial units. Note the gaps in the CBERS data coverage, which were filled with the 500m BUref data.

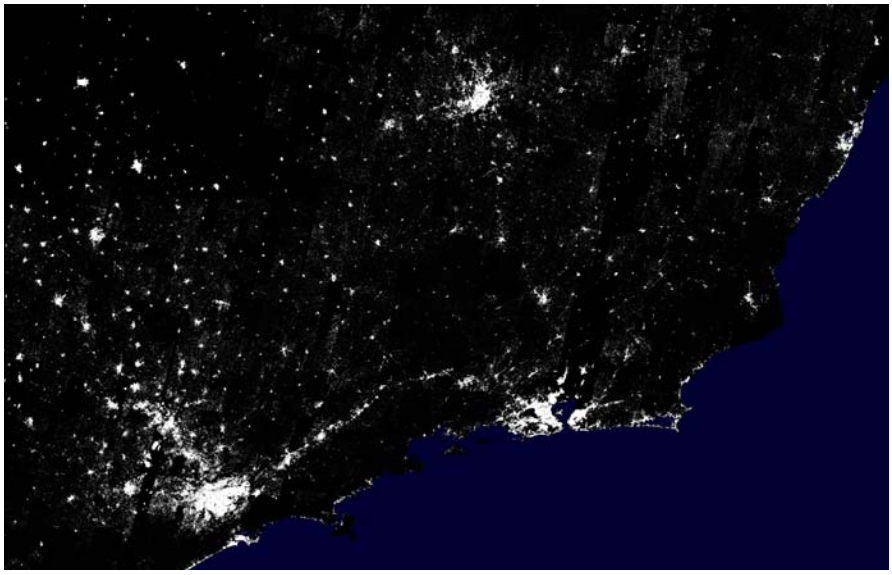


Fig. 11.3 Presence of buildings in the Sao Paolo–Rio de Janeiro–Belo Horizonte triangle. Note that gaps in the CBERS2-B HRC data have been filled with the BUREF layer.

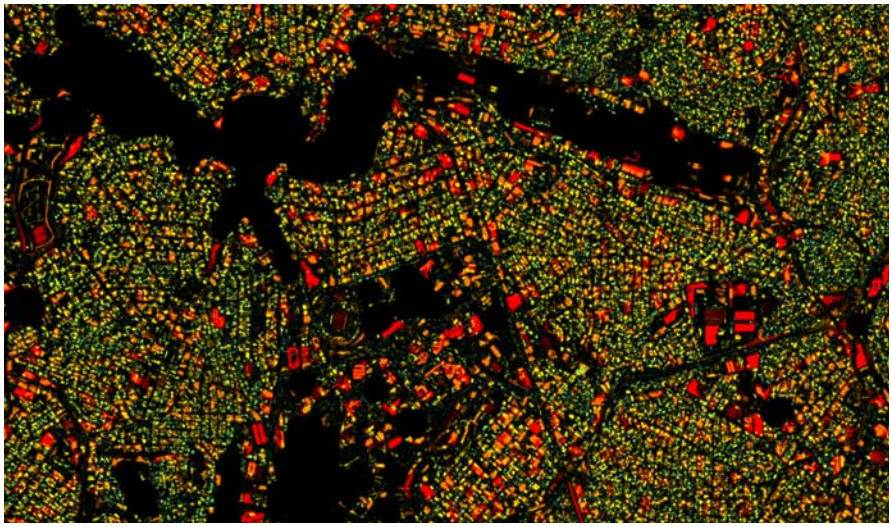


Fig. 11.4 Belo Horizonte (Brazil). Zoom into the area south of the airport. The image shows the 'average building size' layer at a 1:10,000 map scale. The colour-coding follows the blue-green-yellow-red order on increasing size of the built up structures.

11.5 Conclusions

This study confirms that it is possible to use the CBERS-2B data to automatically map the settlements of Brazil. Although only some parts of the country could be analysed until now, the study can be seen as a proof-of-concept with the major terrestrial ecoregions being included and both rural and urbanised areas being mapped. The next steps will be the extension of the processing to all the areas covered by CBERS2-B HRC data. More, research is still needed to validate in a consistent and sustainable way the building size and density information, and to make use of the information produced in the various application areas. This could include the linking of the automatically derived settlement information to other conventionally produced data sets (e.g. the census of Brazil).

Chapter 12
Conclusions

12.1 Summary of the results

This report provides a proof-of-concept for the possibility to build an advanced and up-to-date global human settlement layer (GHSL) derived from HR and VHR optical remotely-sensed data. The test involved 24.3 millions of square kilometres spread over four continents, which were automatically mapped using imagery collected by a variety of optical satellite and airborne sensors with a spatial resolution ranging from 0.5 to 10m. The area mapped was in 2010 inhabited by an estimated population of 1,3 billion according to the LandScan 2010 data set. To the best of our knowledge, it is the largest test of automatic image classification involving such kind of imagery. Several imaging modes were tested including panchromatic, multispectral and pan-sharpened images. A new multi-scale framework was introduced, integrating the automatic image information retrieval with global available geo-information layers derived from other satellite sensors or GIS modelling. For the first time we demonstrated the capacity to extract automatically information from remotely sensed data at detailed scale and the capacity to control the global consistency of the output both spatially and thematically in realistic scenarios. The robustness of the adopted image features was tested globally with a high variety of input data quality including challenging "worst-case" scenarios. New multi-scale morphological and textural image feature compression and optimisation methods were introduced, together with new learning and classification techniques allowing the processing of HR, VHR image data using low-resolution reference data.

12.2 Critical points

Beside some known limitations included in the design of the experiment such as the absence of radiometric descriptors in the discrimination function, the test highlighted some important weaknesses that can be linked to i) the general input data model or ii) the quality control and validation.

The dominant input data model was "per scene" both individual or tiled, with some important exceptions of large mosaics of aggregated aerial frames. In the cases of tiled scenes or image mosaics, the system was creating a virtual image index that was interpreted as a single whole input image source. This "per scene" processing

paradigm seems natural because most of the critical parameters that are necessary to estimate are depending on the per scene image data collection conditions. Even if this is technically correct it may create problems related to i) the arbitrary and very heterogeneous size (mapped surface) of the input scenes and ii) the probability to fall in unfortunate coverage situations, where for example just a little minority of the available image data corresponds to land with searchable information, or no reference data is available. Those facts may create instability in the information extraction process and mislead the learning phases, hence decreasing the consistency of the global mosaic in output. Mitigation measures will include the development of more refined "discovery" learning and classification I2Q modules, taking into account the spatial patterns of the input scenes and their overlap.

Regarding the quality control and validation, the adoption of low-resolution reference layers allowed the development of a fully automatic information extraction chain working for 90% of the cases, but with the risk of destroying the extremely interesting information extraction, where the HR and VHR image data processing was providing unexpected, but reliable results. Mitigation measures will include a priori ranking of the visual inspection according to the results of the automatic matching. In addition, we expect that a systematic collection and system integration of potentially available reference layers at a more detailed scale will provide more reliable reference information. The results of these counter measures will be included in the new release of the GHSL, and will contribute to the assessment of the local reliability of the available low-resolution reference layer. In general, a new Q metric will be developed taking into account i) the mismatch with respect to the available global reference layer, but also ii) the reliability of the reference layers according to the nearest evidences and iii) the known reliability of the same input data with the same operational parameters. This first test helped gaining a deeper insight in these issues.

12.3 Next steps

We are aiming at a continuous improvement of the I2Q system and the GHSL layer with respect tot the processing quality and the areal coverage. The next steps will include i) development and integration of the above described mitigation measures regarding input data model and validation, ii) the introduction of a new radiometric calibration procedure able to work within the constraints of GHSL realistic scenario constraints, iii) the introduction of new radiometric criteria in the I2Q GHSL query, iv) the improvement of the automatic spatial matching module in the I2Q system, which will reduce the errors and the number of non processed scenes (e.g. due to too big displacement or outdated references), v) the introduction of a fully automatic change detection module.

Furthermore, it is expected to extend the IQ GHSL workflow to a wider range of input image data including i) the GMES wall-to-wall complete European coverage of SPOT-5 data of 2012/2013, ii) enlargement of the available areas mapped with CBERS data in Brazil and China, and iii) activation of a process-on-demand mechanism allowing generic users to process any suitable geo-coded image data using the IQ GHSL tools.

Finally, the integration of the GHSL with geo-spatial information extracted from other satellite platforms and active/passive sensors will be studied in the frame of exploitation of cross-platform complementarities.

References

[AA08] H. Akcay and S. Aksoy. "Automatic detection of geospatial objects using multiple hierarchical segmentations". *IEEE Transactions on Geoscience and Remote Sensing* 46.7 (2008), pp. 2097–2111.

[BS72] D. Barnea and H. Silverman. "A class of algorithms for fast digital registration". *IEEE Transactions on Computers* C-21.2 (Nov. 1972), pp. 179–186. DOI: 10.1109/TC.1972.5008923.

[BB05] E. Bartholome and A. Belward. "GLC2000: A new approach to global land cover mapping from Earth observation data". *International Journal of Remote Sensing* 26 (2005), pp. 1959–1977.

[Bau+10] I. Baud, M. Kuffer, K. Pfeffer, and R. Sliuzas. "Understanding heterogeneity in metropolitan India: The added value of remote sensing data for analyzing substandard residential areas". *International Journal of Applied Earth Observation and Geoinformation: JAG* 12.5 (2010), pp. 359–374.

[BPS05] J. Benediktsson, J. Palmason, and J. Sveinsson. "Classification of hyperspectral data from urban areas based on extended morphological profiles". *IEEE Transactions on Geoscience and Remote Sensing* 43.3 (Mar. 2005), pp. 480–491. DOI: 10.1109/TGRS.2004.842478.

[BPK03] J. Benediktsson, M. Pesaresi, and Arnason K. "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations". *IEEE Transactions on Geoscience and Remote Sensing* 9.41 (2003), pp. 1940–1949. DOI: 10.1109/TGRS.2003.814625.

[BGS08] C. Bielski, J. Grazzini, and P. Soille. "Automated morphological image composition for mosaicing large image data sets". In: *Proc. Int. Geosc. and Rem. Sens. Symp.* Barcelona: IEEE Press, July 2008, pp. 4068–4071. DOI: 10.1109/IGARSS.2007.4423743.

[BS05] C. Bielski and P. Soille. "Order independent image compositing". *Lecture Notes in Computer Science* 3617 (Sept. 2005), pp. 1076–1083. DOI: 10.1007/11553595_132.

[Cas+11] E. Castejon, C. Forster, L. Fonseca, and T. Korting. "Orbital image correction for multiple models of geometric transformations". In: *Proc. of Simpósio Brasileiro de Sensoriamento Remoto, 15. (SBSR)*. INPE, São José dos Campos, Brazil, 2011. URL: http://urlib.net/3ERPFQRTRW/39UFJ38.

[CI12] C C.I.E.S.I.N. Center for International Earth Science Information Network Columbia University and C I.A.T. Centro Internacional de Agricultura Tropical. *Gridded Population of the World, Version 3*. June 2012. URL: http://sedac.ciesin.columbia.edu/gpw.

[Cra+12] M. Craglia, K. de Bie, et al. "Digital Earth 2020: Towards the vision for the next decade". *International Journal of Digital Earth* 5.1 (2012), pp. 4–21.

[Cri99] J. Crisp. *Who has counted the refugees?* UNHCR and the politics of numbers. Working Paper No. 12. Policy Research Unit, UNHCR, CP 2500, CH-1211, Geneva, Switzerland. 1999.

[dCE04] J. d'Alge, R. Cartaxo, and G. Erthal. *Geometric quality assessment of CBERS-2*. Tech. rep. Instituto Nacional de Pesquisas Espaciais, 2004. URL: http://mtc-m18.sid.inpe.br/col/dpi.inpe.br/banon/2006/08.03.19.33/doc/appl_13_2004.pdf.

[Dal+10] M. Dalla Mura, J. Benediktsson, B. Waske, and L. Bruzzone. "Morphological attribute profiles for the analysis of very high resolution images". *IEEE Transactions on Geoscience and Remote Sensing* 48.10 (Oct. 2010), pp. 3747–3762.

[Dei+11] U. Deichmann, D. Ehrlich, C. Small, and G. Zeug. *Using high resolution satellite data for the identification of urban natural disaster risk*. World Bank European Union Joint Report. pp 81. Published by World Bank, Washington. 2011.

[Dob+00] J. Dobson, E. Bright, P. Coleman, R. Durfee, and B. Worley. "LandScan: A global population database for estimating populations at risk". *Photogrammetric Engineering and Remote Sensing* 66.7 (July 2000), pp. 849–857.

[EB11] D. Ehrlich and C. Bielski. "Texture based change detection of built-up on SPOT panchromatic imagery using PCA". In: *Proc. of Joint Urban Remote Sensing Event (JURSE)*. 2011, pp. 81–84. DOI: 10.1109/JURSE.2011.5764723.

[Ehr+12] D. Ehrlich, M. Marin Herrera, S. Ferri, T. Kemper, M. Halkia, F. Haag, and M. Pesaresi. *Collection of built-up information from High Resolution satellite imagery*. Tech. rep. [In preparation]. Joint Research Centre of the European Commission, 2012.

[ET12] D. Ehrlich and P. Tenerelli. "Optical satellite imagery for quantifying spatio-temporal dimensions of physical exposure in disaster risk assessments". *Natural Hazards* (2012). DOI: 10.1007/s11069-012-0372-5.

[Elv+01] C. Elvidge, M. Imhoff, K. Baugh, V. Hobson, I. Nelson, and J. Safran. "Nighttime lights of the world: 1994-95". *ISPRS Journal of Photogrammetry and Remote Sensing* 56 (2001), pp. 81–99.

[Gal05] F.J. Gallego. "Stratified sampling of satellite images with a systematic grid of points". *ISPRS Journal of Photogrammetry and Remote Sensing* 59.6 (2005), pp. 369–376. DOI: 10.1016/j.isprsjprs.2005.10.001. URL: http://www.sciencedirect.com/science/article/pii/S0924271605000596.

[Gam+08] P. Gamba, M. Pesaresi, K. Molch, A. Gerhardinger, and G. Lisini. "Anisotropic rotation invariant built-up presence index: Applications

to SAR data". In: *International Geoscience and Remote Sensing Symposium (IGARSS)*. Vol. 5. 1. 2008, pp. V338–V341.

[Goo+12] M. Goodchild, H. Guo, et al. "Next-generation digital earth". *Proceedings of the National Academy of Sciences of the United States of America* 109.28 (2012), pp. 11088–11094.

[Gue+12] L. Gueguen, M. Pesaresi, A. Gerhardinger, and P. Soille. "Characterizing and Counting Roofless Buildings in Very High Resolution Optical Images". *Geoscience and Remote Sensing Letters, IEEE* 9.1 (Jan. 2012), pp. 114–118. ISSN: 1545-598X. DOI: 10.1109/LGRS.2011.2161750.

[GPS11] L. Gueguen, M. Pesaresi, and P. Soille. "An interactive image mining tool handling gigapixel images". In: *Proc. of IGARSS 2011*. IEEE Press, July 2011, pp. 1581–1584. DOI: 10.1109/IGARSS.2011.6049373.

[GSP] L. Gueguen, P. Soille, and M. Pesaresi. "Structure extraction and characterization from differential morphological decomposition". In: *Proc. of the 7th ESA-EUSC-JRC Conference on Image Information Mining*. Ed. by P. Soille, L. Gueguen, S. D'Elia, P.G. Marchetti, L. Colaiacomo, and M. Datcu. Joint Research Centre of the European Commission, pp. 53–57. DOI: 10.2788/69291.

[GSP11] L. Gueguen, P. Soille, and M. Pesaresi. "Change detection based on information measure". *IEEE Transactions on Geoscience and Remote Sensing* 49.11 (Nov. 2011), pp. 4503–4515. DOI: 10.1109/TGRS.2011.2141999.

[GSP12] L. Gueguen, P. Soille, and M. Pesaresi. "A new built-up presence index based on density of corners". In: *Proc. Int. Symp. on Geoscience and Remote Sensing (IGARSS)*. Munich: IEEE, July 2012, pp. 5398–5401. DOI: 10.1109/IGARSS.2012.6352386.

[Guo+09] H. Guo, L. Lu, J. Ma, M. Pesaresi, and F. Yuan. "An improved automatic detection method for earthquake-collapsed buildings from ADS40 image". *Chinese Science Bulletin* (2009). DOI: 10.1007/s11434-009-0461-3.

[Iri+06] R. Irish, J. Barker, S. Goward, and T. Arvidson. "Characterization of the Landsat-7 ETM+ automatic cloud cover assessment (ACCA) algorithm". *Photogrammetric Engineering and Remote Sensing* 72.10 (Oct. 2006), pp. 1179–1188.

[Jon97] R. Jones. "Component trees for image filtering and segmentation". In: *Proc. of IEEE Workshop on Nonlinear Signal and Image Processing*. Ed. by E. Coyle. Mackinac Island, Sept. 1997. URL: http://www.iwaenc.org/proceedings/1997/nsip97/pdf/scan/ns970311.pdf.

[Jon99] R. Jones. "Connected filtering and segmentation using component trees". *Computer Vision and Image Understanding* 75.3 (1999), pp. 215–228. DOI: 10.1006/cviu.1999.0777.

[Kul68] S. Kullback. *Information Theory and Statistics*. New York: Dover, 1968.

[Li+10] L. Li, Z. Li, R. Zhang, J. Ma, and L. Lei. "Collapsed buildings extraction using morphological profiles and texture statistics - A case study in the 5.12 Wenchuan earthquake". In: *Geoscience and Remote Sensing*

Symposium (IGARSS), 2010 IEEE International. July 2010, pp. 2000–2002. DOI: 10.1109/IGARSS.2010.5652333.

[LW06] D. Lu and Q. Weng. "Use of impervious surface in urban land-use classification". *Remote Sensing of Environment* 102.1-2 (2006), pp. 146–160. ISSN: 0034-4257. DOI: 10.1016/j.rse.2006.02.010. URL: http://www.sciencedirect.com/science/article/pii/S0034425706000721.

[Lu+12] L. Lu, Q. Li, L. Jing, H. Guo, and M. Pesaresi. "Classification of CBERS-02B high resolution image using morphological features for urban areas". In: *Earth Observation and Remote Sensing Applications (EORSA), 2012 Second International Workshop on*. June 2012, pp. 16–20. DOI: 10.1109/EORSA.2012.6261126.

[Mar+11] P. Marpu, K.-S. Chen, C.-Y. Chu, and J. Benediktsson. "Spectral-spatial classification of polarimetric SAR data using morphological profiles". In: *2011 3rd International Asia-Pacific Conference on Synthetic Aperture Radar, APSAR 2011*. 2011, pp. 324–326.

[Moo96] T. Moon. "The expectation-maximization algorithm". *IEEE Signal Processing Magazine* 13.6 (1996), pp. 47–60. DOI: 10.1109/79.543975.

[NLM08] S. Niebergall, A. Loew, and W. Mauser. "Integrative assessment of informal settlements using VHR remote sensing data: The Delhi case study". *IEEE Journal of Applied Earth Observation and Remote Sensing* 1.3 (2008), pp. 193–205.

[Ols+01] D. Olson, E. Dinerstein, et al. "Terrestrial ecoregions of the world: a new map of life on Earth". *Bioscience* 51.11 (2001), pp. 933–938.

[OPS12] G. Ouzounis, M. Pesaresi, and P. Soille. "Differential area profiles: decomposition properties and efficient computation". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.8 (Aug. 2012), pp. 1533–1548. DOI: 10.1109/TPAMI.2011.245.

[OS11] G. Ouzounis and P. Soille. "Pattern spectra from partition pyramids and hierarchies". *Lecture Notes in Computer Science* 6671 (2011), pp. 108–119. DOI: 10.1007/978-3-642-21569-8_10.

[OS12] G. Ouzounis and P. Soille. "The Alpha-Tree Algorithm". 25500 EN (Sept. 2012). DOI: 10.2788/48773.

[OSP11] G. Ouzounis, P. Soille, and M. Pesaresi. "Rubble detection from VHR aerial imagery data using differential morphological profiles". In: *Proc. of 34th Int. Symp. on Remote Sensing of the Environment*. Sydney: Int. Center for Remote Sensing of Environment (ICRSE), Apr. 2011. URL: http://www.isprs.org/proceedings/2011/ISRSE-34/211104015Final00015.pdf.

[OSP12] G. Ouzounis, V. Syrris, and M. Pesaresi. "Human settlement analysis based on textural measures and statistical learning". *Selected Tpoics in Applied Earth Observation and Remote Sensing, IEEE Journal of* (2012). in press.

[PP08] E. Pagot and M. Pesaresi. "Systematic Study of the Urban Postconflict Change Classification Performance Using Spectral and Structural Features in a Support Vector Machine". *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 1.2 (June 2008),

pp. 120–128. ISSN: 1939-1404. DOI: 10.1109/JSTARS.2008.2001154.

[PB01] M. Pesaresi and J. Benediktsson. "A new approach for the morphological segmentation of high-resolution satellite imagery". *Geoscience and Remote Sensing, IEEE Transactions on* 39.2 (Feb. 2001), pp. 309–320. ISSN: 0196-2892. DOI: 10.1109/36.905239.

[PE09] M. Pesaresi and D. Ehrlich. "A methodology to quantify built-up structures from optical VHR imagery". In: *Global Mapping of Human Settlement Experiences, Datasets, and Prospects*. Ed. by P. Gamba and M. Herold. CRC Press, 2009. Chap. 3, pp. 27–58. DOI: 10.1201/9781420083408-c3.

[Pes+11] M. Pesaresi, D. Ehrlich, I. Caravaggi, M. Kauffmann, and C. Louvrier. "Towards Global Automatic Built-Up Area Recognition Using Optical VHR Imagery". *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 4.4 (Dec. 2011), pp. 923–934. ISSN: 1939-1404. DOI: 10.1109/JSTARS.2011.2162579.

[PG11] M. Pesaresi and A. Gerhardinger. "Improved textural built-up presence index for automatic recognition of human settlements in arid regions with scattered vegetation". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4.1 (2011), pp. 16–26.

[PGK08] M. Pesaresi, A. Gerhardinger, and F. Kayitakire. "A robust built-up area presence index by anisotropic rotation-invariant textural measure". *Journal of Earth Observation Applications* 1.3 (Sept. 2008), pp. 180–192. DOI: 10.1109/JSTARS.2008.2002869.

[PHO11] M. Pesaresi, M. Halkia, and G. Ouzounis. "Quantitative estimation of settlement density and limits based on textural measurements". In: *2011 Joint Urban Remote Sensing Event, JURSE 2011 - Proceedings*. 2011, pp. 89–92.

[Pes+10] M. Pesaresi, T. Kemper, L. Gueguen, and P. Soille. "Automatic information retrieval from meter and sub-meter resolution satellite image data in support to crisis management". In: *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*. July 2010, pp. 1792–1795. DOI: 10.1109/IGARSS.2010.5653039.

[POG12] M. Pesaresi, G. Ouzounis, and L. Gueguen. "A new compact representation of morphological profiles: report on first massive VHR image processing at the JRC". In: *Proceedings of SPIE 8390*. Ed. by Sylvia S. Shen and Paul E. Lewis. Vol. 8390. Baltimore, USA, Apr. 2012, p. 8390025. DOI: 10.1117/12.920291.

[Pot+09] D. Potere, A. Schneider, S. Angel, and D. Civco. "Mapping urban areas on a global scale: which of the eight maps now available is more accurate?" *International Journal of Remote Sensing* 30.24 (2009), pp. 6531–6558. DOI: 10.1080/01431160903121134.

[RKI73] R. Haralick, K. Shanmugam, and I. Dinstein. "Textural features for image classification". *IEEE Transactions on Systems, Man, and Cybernetics* 3.6 (Nov. 1973), pp. 610–621. DOI: 10.1109/TSMC.1973.4309314s.

[RSB93] J.-F. Rivest, P. Soille, and S. Beucher. "Morphological gradients". *Journal of Electronic Imaging* 2.4 (Oct. 1993), pp. 326–336. DOI: 10.1117/12.159642.

[SOG98] P. Salembier, A. Oliveras, and L. Garrido. “Antiextensive connected operators for image and sequence processing”. *IEEE Transactions on Image Processing* 7.4 (Apr. 1998), pp. 555–570. DOI: 10.1109/83.663500.

[SS95] P. Salembier and J. Serra. “Flat zones filtering, connected operators, and filters by reconstruction”. *IEEE Transactions on Image Processing* 4.8 (1995), pp. 1153–1160.

[Sat07] D. Satterthwaite. *The transition to a predominantly urban world and its underpinnings*. Human Settlements Working Paper. International Institute for Environment and Development, Sept. 2007. ISBN: 978-1-84369-670-4.

[SFP10] A. Schneider, M. Friedl, and D. Potere. “Monitoring urban areas globally using MODIS 500m data: new methods based on urban ecoregions”. *Remote Sensing of Environment* 114.8 (Aug. 2010), pp. 1733–1746. DOI: 10.1016/j.rse.2010.03.003.

[Ser82] J. Serra. *Image Analysis and Mathematical Morphology*. London: Academic Press, 1982.

[Shy+06] C.-R. Shyu, M. Klaric, G. Scott, and W. Mahamaneerat. “Knowledge Discovery by Mining Association Rules and Temporal-Spatial Information from Large-Scale Geospatial Image Databases”. In: *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on*. Aug. 2006, pp. 17–20. DOI: 10.1109/IGARSS.2006.9.

[Soi03] P. Soille. *Morphological Image Analysis: Principles and Applications*. 2nd. Berlin Heidelberg New York: Springer-Verlag, 2003.

[Soi06] P. Soille. “Morphological image compositing”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.5 (May 2006), pp. 673–683. DOI: 10.1109/TPAMI.2006.99. URL: http://dx.doi.org/10.1109/TPAMI.2006.99.

[Soi08] P. Soille, ed. *The IMAGE-2006 Mosaic Project*. Joint Research Centre of the European Commission, 2008. DOI: 10.2788/25572.

[SN12] P. Soille and L. Najman. “On morphological hierarchical representations for image processing and spatial data clustering”. *Lecture Notes in Computer Science* 7346 (2012), pp. 43–67. DOI: 10.1007/978-3-642-32313-3_4.

[TH86] Q. Tian and M. Huhns. “Algorithms for subpixel registration”. *Computer Vision, Graphics, and Image Processing* 35.2 (1986), pp. 220–233.

[TAP11a] G. Trianni, E. Angiuli, and M. Pesaresi. “Improved anisotropic rotation invariant built-up presence index for a global human settlements layer (GHSL from medium resolution SAR data”. In: *The seventh conference on image information mining: Geospatial intelligence from earth observation (IIM), 2011 ESA EUSC JRC*. June 2011.

[TAP11b] G. Trianni, E. Angiuli, and M. Pesaresi. “Statistical analysis of anisotropic rotation-invariant textural measurements of human settlements from multitemporal SAR data”. In: *Urban Remote Sensing Event (JURSE), 2011 Joint*. Apr. 2011, pp. 117–120. DOI: 10.1109/JURSE.2011.5764733.

[TGD04] C. Tucker, D. Grant, and J. Dykstra. “NASA’s Global orthorectified Landsat data set”. *Photogrammetric Engineering and Remote Sensing* 70.3 (Mar. 2004), pp. 313–322.

[UN 12] UN Department of Economic & Social Affairs Population Division. *World urbanization prospects: the 2011 revision*. Apr. 2012. URL: `http://esa.un.org/unpd/wup/index.htm`.

[Wan+04] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. “Image quality assessment: from error visibility to structural similarity”. *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612.

European Commission
EUR 25662 EN– Joint Research Centre – Institute for the Protection and Security of the Citizen

Title: **A Global Human Settlement Layer from Optical High Resolution Imagery**

Authors: Martino Pesaresi, Xavier Blaes, Daniele Ehrlich, Stefano Ferri, Lionel Gueguen,, Fernand Haag, Matina Halkia, Johannes Heinzel, Mayeul Kauffmann, Thomas Kemper, Georgios K. Ouzounis, Marco Scavazzon, Pierre Soille, Vasileios Syrris, and Luigi Zanchetta

Luxembourg: Publications Office of the European Union

2012 – 121 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online), ISSN 1018-5593 (print),

ISBN 978-92-79-27988-1 (pdf)
ISBN 978-92-79-27989-8 (print)

doi:10.2788/73897

Abstract

A general framework for processing of high and very-high resolution imagery for creating a Global Human Settlement Layer (GHSL) is presented together with a discussion on the results of the first operational test of the production workflow. The test involved the mapping of 24.3 millions of square kilometres of the Earth surface spread over four continents, corresponding to an estimated population of 1.3 billion of people in 2010. The resolution of the input image data ranges from 0.5 to 10 meters, collected by a heterogeneous set of platforms including satellite SPOT (2 and 5), CBERS-2B, RapidEye (2 and 4), WorldView (1 and 2), GeoEye-1, QuickBird-2, Ikonos-2, and airborne sensors. Several imaging modes were tested including panchromatic, multispectral and pan-sharpened images. A new fully automatic image information extraction, generalization and mosaic workflow is presented that is based on multiscale textural and morphological image features extraction. New image feature compression and optimization are introduced, together with new learning and classification techniques allowing for the processing of HR/VHR image data using low-resolution thematic layers as reference. A new systematic approach for quality control and validation allowing global spatial and thematic consistency checking is proposed and applied. The quality of the results is discussed by sensor, by band, by resolution, and eco-regions. Critical points, lessons learned and next steps are highlighted

LB-NA-25662-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

ISBN 978-92-79-27988-1

Publications Office